

DOCUMENT RETRIEVAL REQUEST FORM

Requester's Name: HOCUMES Case Serial Number: 09/22/05 Art Unit/Org.: 212
 Phone: 308628 ~~390-1222~~ Building: PK2 Room Number: 200
 Class/Sub-Class: _____
 Date of Request: 11/04/03 Date Needed By: 11/04/03
 Paste or add text of citation or bibliography: Paste Citation Only one request per form. Original copy only. ☐

Author/Editor: _____
 Journal/Book Title: _____
 Article Title: _____
 Volume Number: _____ Report Number: _____ Pages: _____
 Issue Number: _____ Series Number: _____ Year of Publication: _____
 Publisher: _____
 Remarks: (18) SEE ATTACHED
470745

STANDARD ONLY

Monthly Accession Number: _____

Library Action	PTO		LC		NAL		NIH		NLM		NIST		Other	
	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd
Local Attempts	✓													
Date	<u>11/4</u>													
Initials	<u>me</u>													
Results	<u>11</u>													
Examiner Called														
Page Count														
Money Spent														

Source: _____ Date: _____
 Remarks/Comments: CRST ordered
Complete
Fixed
 Ordered From: _____
 Comments: _____
 O/N - Under NLM means Overnight

Document

Select the documents you wish to save or order by clicking the box next to the document, or click the link above the document to order directly.

18

locally as: ☐ include search strategy

☐ document 22 of 117 [Order Document](#)

INSPEC - 1969 to date (INZZ)

Accession number & update

6650206, C2000-08-7250R-045; 20000701.

Title

Recognizing structure in Web pages using similarity queries.

Author(s)

Cohen-W-W.

Author affiliation

Shannon Lab, AT&T Bell Labs, Florham Park, NJ, USA.

Source

Proceedings Sixteenth National Conference on Artificial Intelligence (AAAI-99). Eleventh Innovative Applications of Artificial Intelligence Conference (IAAI-99), Orlando, FL, USA, 18-22 July 1999. Sponsors: American Assoc. Artificial Intelligence, ACM/SIGART, Defense Advance Res. Projects Agency, et al.
In: p.59-66, 1999.

ISSN

ISBN: 0-262-51106-1.

Publication year

1999.

Language

EN.

Publication type

CPP, Conference Paper.

Treatment codes

P Practical.

Abstract

We present general-purpose methods for recognizing certain types of structure in HTML documents. The methods are implemented using WHIRL, a "soft" logic that incorporates a notion of textual similarity developed in the information retrieval community. In an experimental evaluation on 82 Web pages, the structure ranked first by our method is "meaningful", i.e., a structure that was used in a hand-coded "wrapper", or extraction program, for the page, nearly 70% of the time. This improves on a value of 50% obtained by an earlier method. With appropriate background information, the structure-recognition methods described can also be used to learn a wrapper from examples, or for maintaining a wrapper as a Web page changes format. In these settings, the top-ranked structure is meaningful nearly 85% of the time. (20 refs).

Descriptors

data-structures; formal-logic; information-resources; Internet; learning-systems; pattern-recognition; query-processing.

BEST AVAILABLE COPY

Recognizing Structure in Web Pages using Similarity Queries

William W. Cohen

AT&T Labs-Research Shannon Laboratory
180 Park Avenue Florham Park, NJ 07932
wcohen@research.att.com

Abstract

We present general-purpose methods for recognizing certain types of structure in HTML documents. The methods are implemented using WHIRL, a "soft" logic that incorporates a notion of textual similarity developed in the information retrieval community. In an experimental evaluation on 82 Web pages, the structure ranked first by our method is "meaningful"—i.e., a structure that was used in a hand-coded "wrapper", or extraction program, for the page—nearly 70% of the time. This improves on a value of 50% obtained by an earlier method. With appropriate background information, the structure-recognition methods we describe can also be used to learn a wrapper from examples, or for maintaining a wrapper as a Web page changes format. In these settings, the top-ranked structure is meaningful nearly 85% of the time.

Introduction

Web-based information integration systems allow a user to query structured information that has been extracted from the Web (Levy, Rajaraman, & Ordille 1996; Garcia-Molina *et al.* 1995; Knoblock *et al.* 1998; Genesereth, Keller, & Dushka 1997; Lacroix, Sahuguet, & Chandrasekar 1998; Mecca *et al.* 1998; Tomasic *et al.* 1997). In most such systems, a different *wrapper* must be written for each Web site that is accessed. A *wrapper* is a special-purpose program that extracts information from Web pages written in a specific format. Because data can be presented in many different formats, and because Web pages frequently change, building and maintaining wrappers is time-consuming and tedious. To reduce the cost of building wrappers, some researchers have proposed special languages for writing wrappers (Hammer *et al.* 1997; Cohen 1998b), or semi-automated tools for wrapper construction (Ashish & Knoblock 1997). Others have implemented systems that allow wrappers to be trained from examples (Kushmerick, Weld, & Doorenbos 1997; Hsu 1998; Muslea, Minton, & Knoblock 1998). Data exchange standards like XML have also been proposed, although as yet none are in widespread use.

Here, we explore another approach to this problem: developing general-purpose methods for automatically

Exploding porpoises, over four score and seven, well before configuration.

- *Department of Computer and Information Sciences, University of New Jersey.* Citrus flavorings: green, marne, clean and under lien.
- *Computer Engineering Center, Lough Polytechnical Institute.* This, that page extensionally left to rights of manatees.
- *Electrical Engineering and Computer Science Dept, Bismark State College.* Tertiary; where cola substitutes are frequently underutilized.

This page under construction. (Last update: 9/23/98.)

Figure 1: Nonsense text with a meaningful structure.

recognizing structure in HTML documents. Our ultimate goal is to extract structured information from Web pages without any page-specific programming or training.

To motivate this approach, consider Figure 1. To a human reader, this text is perceived as containing a list of three items, each containing the italicized name of a university department, with the university name underlined. This apparently meaningful structure is recognized without previous knowledge or training, even though the text is ungrammatical nonsense and the university names are imaginary. This suggests that people employ general-purpose, page-independent strategies for recognizing structure in documents. Incorporating similar strategies into a system that automatically (or semi-automatically) constructs wrappers would clearly be valuable.

Below we show that effective structure recognition methods for certain restricted types of list structures can be encoded compactly and naturally, given appropriate tools. In particular, we will present several methods that can be concisely implemented in WHIRL (Cohen 1998a), a "soft" logic that includes both "soft" universal quantification, and a notion of textual similarity developed in the information retrieval (IR) community. The structure-recognition methods we present are based on natural heuristics, such as detecting repetition of sequences of markup commands, and detecting repeated patterns of "familiar-looking" strings.

The methods can be used in a page-independent manner: given an HTML page, but no additional information about it, the methods produce a ranked list of pro-

HTML source for a simple list:

```
<html><head>...</head>
<body>
<h1>Editorial Board Members</h1>
<table> <tr>
  <td>G. R. Emlin, Lucent</td>
  <td>Harry Q. Bovik, Cranberry U</td></tr>
<tr>
  <td>Bat Gangley, UC/Bovine</td>
  <td>Pheobe L. Mind, Lough Tech</td>
...

```

Extracted data:

G. R. Emlin, Lucent
Harry Q. Bovik, Cranberry U
...

HTML source for a simple hotlist:

```
<html><head>...</head>
<body><h1>Publications for Pheobe Mind</h1>
<ul>
<li>Optimization of fuzzy neural networks using
distributed parallel case-based genetic knowledge discovery
  (<a href="buzz.pdf">PDF</a>)</li>
<li>A linear-time version of GSAT
  (<a href="peqnp.ps">postscript</a>)</li>
...

```

Extracted data:

Optimization ... (PDF)	buzz.pdf
A linear-time version of ...	peqnp.ps
...	...

Figure 2: A simple list, a simple hotlist, and the data that would be extracted from each.

posed "structures" found in the page. This ranking is generally quite useful: in an experimental evaluation on 82 Web pages associated with real extraction problems, the top-ranked structure is "meaningful" (as defined below) nearly 70% of the time. This improves on an earlier method (Cohen & Fan 1999), which proposes meaningful structures about 50% of the time on the same data.

By providing different types of additional information, about a page, the same methods can also be used for page-specific wrapper learning as proposed by Kushmerick *et al* (1997), or for updating a wrapper after the format of a wrapped page has changed. When used for page-specific learning or wrapper update, the top-ranked structure is meaningful nearly 85% of the time.

Background

Benchmark problems

We begin by clarifying the structure-recognition problem, with the aim of stating a task precise enough to allow quantitative evaluation of performance. Deferring for now the question of what a "structure" is, we propose to rate the "structures" identified by our methods as either *meaningful* or *not meaningful*. Ideally, a structure in a Web page would be rated as *meaningful* iff it contains structured information that could plausibly be extracted from the page. Concretely, in our experiments, we will use pages that were actually wrapped by an information integration system, and consider a structure as meaningful iff it corresponds to information actually extracted by an existing, hand-coded wrapper for that page.

In this paper, we will restrict ourselves to wrappers in two narrow classes (and therefore, to a narrow class of potential structures). We call these wrapper classes *simple lists* and *simple hotlists*. In a page containing a *simple list*, the information extracted is a one-column relation containing a set of strings s_1, \dots, s_N , and each s_i is all the text that falls below some node n_i in the HTML parse tree for the page. In a *simple hotlist*, the extracted information is a two-column relation, contain-

ing a set of pairs $\langle s_1, u_1 \rangle, \dots, \langle s_n, u_n \rangle$; each s_i is all the text that falls below some node n_i in the HTML parse tree; and each u_i is a URL that is associated with some HTML anchor element a_i that appears somewhere inside n_i . Figure 2 shows the HTML source for a simple list and a simple hotlist, and the data that is extracted from each.

This restriction is based on our experience with a working information integration system (Cohen 1998b). Of 111 different wrapper programs written for this system, 82 (or nearly 75%) were based on simple lists or simple hotlists, as defined above.¹ We will use this corpus of problems in the experiments described below.

The vector space representation for text

Our ability to perceive structure in the text of Figure 1 is arguably enhanced by the regular appearance of substrings that are recognizable as (fictitious) university names. These strings are recognizable because they "look like" the names of real universities. Implementing such heuristics requires a precise notion of similarity for text, and one such notion is provided by the *vector space model* of text.

In the vector space model, a piece of text is represented as a *document vector* (Salton 1989). We assume a vocabulary T of terms; in this paper, terms are word stems produced by the Porter stemming algorithm (Porter 1980). A *document vector* is a vector of real numbers $\vec{v} \in \mathcal{R}^{|T|}$, each component of which corresponds to a term $t \in T$. We will denote the component of \vec{v} which corresponds to $t \in T$ by v_t , and employ the TF-IDF weighting scheme (Salton 1989): for a document vector \vec{v} appearing in a collection C , we let v_t be zero if the term t does not occur in text represented by \vec{v} , and otherwise let $v_t = (\log(TF_{\vec{v},t}) + 1) \cdot \log(IDF_t)$. In this formula, $TF_{\vec{v},t}$ is the number of times that

¹We say "based on" because some lists also included pre-processing or filtering steps. We note also that the relative simplicity of wrappers is due in part to special properties of the information integration system. Further discussion of this dataset can be found elsewhere (Cohen & Fan 1999).

term t occurs in the document represented by \vec{v} , and $IDF_t = \frac{|C|}{|C_t|}$, where C_t is the set of documents in C that contain t .

In the vector space model, the similarity of two document vectors \vec{v} and \vec{w} is given by the formula $SIM(\vec{v}, \vec{w}) = \sum_{t \in T} \frac{v_t w_t}{\|\vec{v}\| \|\vec{w}\|}$. Notice that $SIM(\vec{v}, \vec{w})$ is always between zero and one, and that similarity is large only when the two vectors share many "important" (highly weighted) terms.

The WHIRL logic

Overview. WHIRL is a logic in which the fundamental items that are manipulated are not atomic values, but entities that correspond to fragments of text. Each fragment is represented internally as a document vector, as defined above; this means the similarity between any two items can be computed. In brief, WHIRL is non-recursive, function-free Prolog, with the addition of a built-in similarity predicate; rather than being true or false, a similarity literal is associated with a real-valued "score" between 0 and 1; and scores are combined as if they were independent probabilities.

As an example of a WHIRL query, let us suppose that the information extracted from the simple list of Figure 2 is stored as a predicate $ed_board(X)$. Suppose also that the information extracted from the hotlist of Figure 2, together with a number of similar bibliography hotlists, has been stored in a predicate $paper(Y, Z, U)$, where Y is an author name, Z a paper title, and U a paper URL. For instance, the following facts may have been extracted and stored: $ed_board("Pheobe L. Mind, Lough Tech")$, and $paper("Pheobe Mind", "A linear-time version of GSAT", "http://.../pegnp.ps")$. Using WHIRL's similarity predicate " \sim ", the following query might be used to find papers written by editorial board members:

$\leftarrow ed_board(X) \wedge paper(Y, Z, U) \wedge X \sim Y$

The answer to this query would be a list of substitutions θ , each with an associated score. Substitutions that bind X and Y to similar documents would be scored higher. One high-scoring substitution might bind X to "Pheobe L. Mind, Lough Tech" and Y to "Pheobe Mind".

Below we will give a formal summary of WHIRL. A complete description is given elsewhere (Cohen 1998a).

WHIRL semantics. Like a conventional deductive database (DDB) program, a WHIRL program consists of two parts: an *extensional database* (EDB), and an *intensional database* (IDB). The IDB is a non-recursive set of function-free definite clauses. The EDB is a collection of ground atomic facts, each associated with a numeric *score* in the range $[0, 1]$. In addition to the types of literals normally allowed in a DDB, clauses in the IDB can also contain *similarity literals* of the form $X \sim Y$, where X and Y are variables. A WHIRL predicate definition is called a *view*. We will assume below views are *flat*—that is, that each clause body in the

view contains only literals associated with predicates defined in the EDB. Since WHIRL does not support recursion, views that are not flat can be "flattened" (unfolded) by repeated resolution.

In a conventional DDB, the answer to a conjunctive query would be the set of ground substitutions that make the query true. In WHIRL, the notion of provability will be replaced with a "soft" notion of *score*, which we will now define. Let θ be a ground substitution for B . If $B = p(X_1, \dots, X_n)$ corresponds to a predicate defined in the EDB, then $SCORE(B, \theta) = s$ if $B\theta$ is a fact in the EDB with score s , and $SCORE(B, \theta) = 0$ otherwise. If B is a similarity literal $X \sim Y$, then $SCORE(B, \theta) = SIM(\vec{x}, \vec{y})$, where $\vec{x} = X\theta$ and $\vec{y} = Y\theta$. If $B = B_1 \wedge \dots \wedge B_k$ is a conjunction of literals, then $SCORE(B, \theta) = \prod_{i=1}^k SCORE(B_i, \theta)$. Finally, consider a WHIRL view, defined as a set of clauses of the form $A_i \leftarrow Body_i$. For a ground atom a that is an instance of one or more A_i 's, we define the *support* of a , $SUPPORT(a)$, to be the set of all pairs $(\sigma, Body_i)$ such that $A_i\sigma = a$, $Body_i\sigma$ is ground, and $SCORE(Body_i, \sigma) > 0$. We define the score of an atom a (for this view) to be

$$1 - \prod_{(\sigma, Body_i) \in SUPPORT(a)} (1 - SCORE(Body_i, \sigma))$$

This definition follows from the usual semantics of logic programs, together with the observation that if e_1 and e_2 are independent events, then $Prob(e_1 \vee e_2) = 1 - (1 - Prob(e_1))(1 - Prob(e_2))$.

The operations most commonly performed in WHIRL are to define and *materialize* views. To materialize a view, WHIRL finds a set of ground atoms a with non-zero score s_a for that view, and adds them to the EDB. Since in most cases, only high-scoring answers will be of interest, the materialization operator takes two parameters: τ , an upper bound on the number of answers that are generated, and ϵ , a lower bound on the score of answers that are generated.

Although the procedure used for combining scores in WHIRL is naive, inference in WHIRL can be implemented quite efficiently. This is particularly true if ϵ is large or τ is small, and if certain approximations are allowed (Cohen 1998a).

The "many" construct. The structure-recognition methods we will present require a recent extension to the WHIRL logic: a "soft" version of universal quantification. This operator is written $many(Template, Test)$ where the *Test* is an ordinary conjunction of literals, and the *Template* is a single literal of the form $p(Y_1, \dots, Y_n)$, where p is an EDB predicate and the Y_i 's are all distinct; also, the Y_i 's may appear only in *Test*. The score of a "many" clause is the weighted average score of the *Test* conjunction on items that match the *Template*. More formally, for a substitution θ and a conjunction W ,

$$SCORE(many(p(Y_1, \dots, Y_n), Test), \theta) =$$

$$\sum_{(s, a_1, \dots, a_k) \in P} \frac{s}{S} \cdot \text{SCORE}(\text{Test}, (\theta \circ \{Y_i = a_i\}_i))$$

where P is the set of all tuples (s, a_1, \dots, a_k) such that $p(a_1, \dots, a_k)$ is a fact in the EDB with score s ; S is the sum of all such scores s ; and $\{Y_i = a_i\}_i$ denotes the substitution $\{Y_1 = a_1, \dots, Y_k = a_k\}$.

As an example, the following WHIRL query is a request for editorial board members that have written "many" papers on neural networks.

```
q(X) ← ed.board(X) ∧
      many(papers(Y, Z, W),
           (X ~ Y ∧ Z ~ "neural networks"))
```

Recognizing structure with WHIRL

Encoding HTML pages and wrappers

We will now give a detailed description of how structure-recognition methods can be encoded in WHIRL. We begin with a description of the encoding used for an HTML page.

To encode an HTML page in WHIRL, the page is first parsed. The HTML parse tree is then represented with the following EDB predicates.

- $\text{elt}(\text{Id}, \text{Tag}, \text{Text}, \text{Position})$ is true if Id is the identifier for a parse tree node, Tag is the HTML tag associated with n , Text is all of the text appearing in the subtree rooted at n , and Position is the sequence of tags encountered in traversing the path from the root to n . The value of Position is encoded as a document containing a single term t_{pos} , which represents the sequence, e.g., $t_{\text{pos}} = \text{"html_body_ul_li"}$.
- $\text{attr}(\text{Id}, \text{AName}, \text{AValue})$ is true if Id is the identifier for node n , AName is the name of an HTML attribute associated with n , and AValue is the value of that attribute.
- $\text{path}(\text{FromId}, \text{ToId}, \text{Tags})$ is true if Tags is the sequence of HTML tags encountered on the path between nodes FromId and ToId . This path includes both endpoints, and is defined if $\text{FromId} = \text{ToId}$.

As an example, wrappers for the pages in Figure 2 can be written using these predicates as follows.

```
page1(NameAffil) ←
  elt(., ., NameAffil, "html\_body\_table\_tr\_td").
page2(Title, Url) ←
  elt(ContextElt, ., Title, "html\_body\_ul\_li")
  ∧ path(ContextElt, AnchorElt, "li\_a")
  ∧ attr(AnchorElt, "href", Url).
```

Next, we need to introduce an appropriate encoding of "structures" (and in so doing, make this notion precise.) Most simple lists and hotlists in our benchmark collection can be wrapped with some variant of either the *page1* or *page2* view, in which the constant strings (e.g., "html_body_ul_li" and "li_a") are replaced with different values. Many of the remaining pages can be wrapped by views consisting of a disjunction of such clauses.

We thus introduce a new construct to formally represent the informal idea of a "structure" in a structured document: a *wrapper piece*. In the most general setting, a *wrapper piece* consists of a *clause template* (e.g., a generic version of *page2* above), and a set of *template parameters* (e.g., the pair of constants "html_body_ul_li" and "li_a"). In the experiments below, we consider only two clause templates—the ones suggested by the examples above—and also assume that the recognizer knows, for each page, if it should look for list structures or hotlist structures. In this case, the clause template need not be explicitly represented; a wrapper piece for a *page2* variant can be represented simply as a pair of constants (e.g., "html_body_ul_li" and "li_a"), and a wrapper piece for a *page1* variant can be represented as a single constant (e.g., *html_body_table_tr_td*).

For brevity, we will confine the discussion below to methods that recognize simple hotlist structures analogous to *page2*, and will assume that structures are encoded by a pair of constants *Path1* and *Path2*. However, most of the methods we will present have direct analogs that recognize simple lists.

Enumerating and ranking wrappers

We will now describe three structure-recognition methods based on these encodings. We begin with some basic building blocks. Assuming that some page of interest has been encoded in WHIRL's EDB, materializing the WHIRL view *possible_piece*, shown in Figure 3, will generate all wrapper pieces that would extract at least one item from the page. The *extracted_by* view determines which items are extracted by each wrapper piece, and hence acts as an interpreter for wrapper pieces.

Using these views in conjunction with WHIRL's soft universal quantification, one can compactly state a number of plausible recognition heuristics. One heuristic is to prefer wrapper pieces that extract many items; this trivial but useful heuristic is encoded in the *fruitful_piece* view. Recall that materializing a WHIRL view results in a set of new atoms, each with an associated score. The *fruitful_piece* view can thus be used to generate a ranked list of proposed "structures" by simply presenting all *fruitful_piece* facts to the user in decreasing order by score.

Another structure-recognition method is suggested by the observation that in most hotlists, the text associated with the anchor is a good description of the associated object. This suggests the *anchorlike_piece* view, which adds to the *fruitful_piece* view an additional "soft" requirement that the text *Text1* extracted by the wrapper piece be similar to the text *Text2* associated with the anchor element.

A final structure-recognition method is shown in the Figure as the *R_like_piece* view. This view is a copy of *fruitful_piece* in which the requirement that many items are extracted is replaced by a requirement that many "R like" items are extracted, where an item is "R like" if it is similar to some second item X that is stored in the EDB relation R . The "soft" semantics of the *many*

```

fruitful.piece(Path1,Path2) ←
possible.piece(Path1,Path2) ∧
many( extracted.by(Path1a,Path2a,-,-),
(Path1a=Path1 ∧ Path2a=Path2) ).
possible.piece(Path1,Path2) ←
elt(TextElt, -, -, Path1)
∧ elt(AnchorElt, -, "a", -)
∧ attr(AnchorElt, "href", -)
∧ path(TextElt, AnchorElt, Path2).
extracted.by(Path1,Path2,TextElt,AnchorElt) ←
elt(TextElt, -, -, Path1)
∧ path(TextElt, AnchorElt, Path2).

```

```

anchorlike.piece(Path1,Path2) ←
possible.piece(Path1,Path2) ∧
many( extracted.by(Path1a,Path2a,TElt,AElt),
(Path1a=Path1 ∧ Path2a=Path2
∧ elt(TElt,-,Text1,-) ∧ elt(AElt,-,Text2,-) ∧ Text1~Text2 ) ).
R.like.piece(Path1,Path2) ←
possible.piece(Path1,Path2) ∧
many( R.extracted.by(Path1a,Path2a,-,-),
(Path1a=Path1 ∧ Path2a=Path2) ).
R.extracted.by(Path1,Path2,TextElt,AnchorElt) ←
elt(TextElt, -, Text, Path1)
∧ path(TextElt, AnchorElt, Path2)
∧ R(X) ∧ Text~X.

```

Figure 3: WHIRL programs for recognizing plausible structures in an HTML page. (See text for explanation.)

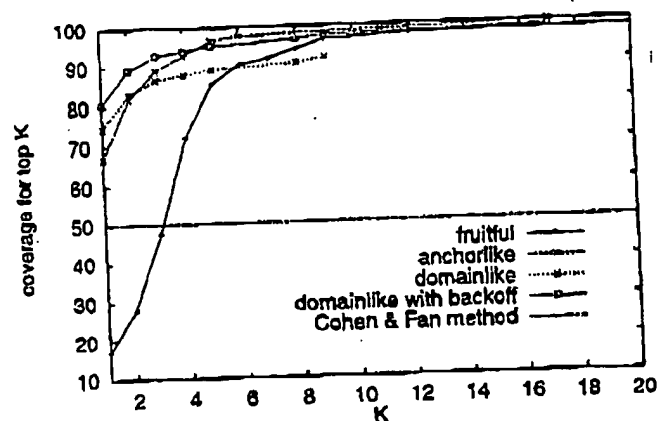


Figure 4: Performance of ranking heuristics that use little or no page-specific information.

construct imply that more credit is given to extracting items that match an item in R closely, and less credit is given for weaker matches. As an example, suppose that R contains a list of all accredited universities in the US. In this case, the $R_{like}.piece$ would prefer wrapper pieces that extract many items that are similar to some known university name; this might be useful in processing pages like the one shown in Figure 1.

Experiments

Ranking programs with (almost) no page-specific information

We will now evaluate the three structure-recognition methods shown in Figure 3. We took the set of 82 hand-coded list and hotlist wrappers described above, and paired each hand-coded wrapper with a single Web page that was correctly wrapped. We then analyzed the hand-coded wrapper programs, and determined which wrapper pieces they contained. The result of this preprocessing was a list of 82 Web pages, each of which is associated with a set of "meaningful" wrapper pieces. To evaluate a method, we materialize the appropriate

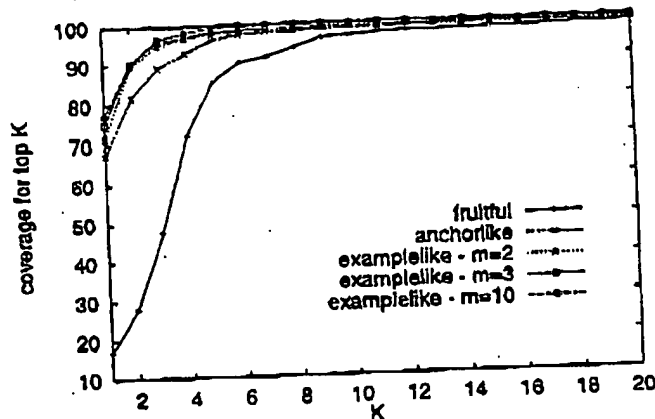


Figure 5: Performance of ranking heuristics that use page-specific training examples.

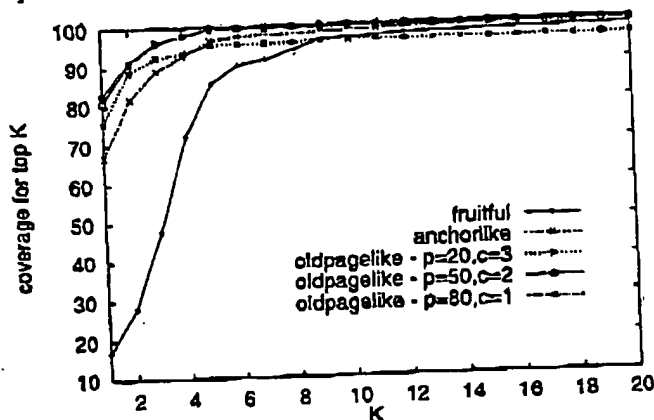


Figure 6: Performance of ranking heuristics that use text extracted from an previous version of the page.

view,² thus generating a ranked list of proposed struc-

²Thresholds of $\epsilon = 0.5$ and $r = 100,000$ were used in materializing the view $R_{extracted.by}$, and thresholds of $\epsilon = 0$ and $r = 100,000$ were used elsewhere. We also assume that the system "knows" whether a list or a hotlist is to be extracted from each page: i.e., we apply to each hotlist page

tures. A good method is one that ranks meaningful structures ahead of non-meaningful structures.

To obtain useful aggregate measures of performance, it is useful to consider how a structure-recognition method might be used. One possibility is *interactive use*: given a page to wrap, the method proposes wrapper pieces to a human user, who then examines them in order, and manually selects pieces to include in a wrapper for the page. To evaluate performance, we vary K and record for each K the *coverage* at rank K —that is, the percentage of the 82 problems that can be wrapped using pieces ranked in the top K . The resulting “coverage curve” can be plotted. We also compute a measure we call the *average number of skips*: the average number of non-meaningful pieces that are ranked ahead of some meaningful piece; or, equivalently, the average number of wrapper pieces that would be unnecessarily examined (skipped over) by the user.

Another possibility use for the system is *batch use*: given a page, the method proposes a single structure, which is then used by a calling program without any filtering. For batch use, a natural measure of performance is the percentage of the time that the top-ranked structure is meaningful. Below, we call this measure *accuracy at rank 1*, and define *error rate at rank 1* analogously.³

Figure 4 shows the coverage curves obtained from methods that require no page-specific information. For comparison, we also show the performance of a earlier structure-recognition method (Cohen & Fan 1999). (To summarize this method briefly, structure recognition is reduced to the problem of classifying nodes in an HTML parse tree as to whether or not they are contained in some meaningful structure. The node-classification problem can then be solved by off-the-shelf inductive learning methods such as CART (Breiman et al. 1984) or RIPPER (Cohen 1995).) This method produces a single wrapper program (which may correspond to multiple wrapper pieces), rather than a ranked list of wrapper pieces. On the data used here, the wrapper proposed coincides with the true wrapper, or some close approximation of it, on exactly half the cases.

The *anchorlike* method⁴ performs quite well, obtaining accuracy at rank 1 of nearly 70%, and an average of 0.9 skips. (These numbers are summarized in Table 1). Even the strawman *fruitful* method works surprisingly well in interactive use, obtaining an average number of skips of only 3.3; however, for batch use, its accuracy

only structure-recognition views that recognize hotlists, and apply to each list page only views that recognize lists.

³Note that accuracy at rank 1 is not identical to coverage at $K = 1$; the former records the number of times the top-ranked wrapper piece is part of the target wrapper, and the latter records number of times the top-ranked wrapper piece is the *only* piece in the target wrapper.

⁴Recall that the *anchorlike* can only be applied to hotlists. In the curve labeled *anchorlike*, we used the *fruitful* method for simple list wrappers, and the *anchorlike* method for simple hotlist wrappers.

at rank 1 is less than 20%.

The third curve shown in Figure 4, labeled *domainlike*, is an instance of the *R_{like}-piece* method in which R contains a large list of items in the same domain as the items to be extracted. (For instance, if the data to be extracted is a list of universities, then R would be a second list of universities.) We consider this structure-recognition method in this section because, although it does require some page-specific information, the information required is quite easy to obtain.⁵ The average skip rate and error at rank 1 for the *domainlike* method are roughly half that of *anchorlike*. However, this method does not obtain 100% coverage, as in some fraction of the problems, the secondary relation R is either unavailable or misleading.

The final curve in Figure 4, labeled “*domainlike with backoff*”, is a simple combination of the *domainlike* and *anchorlike* strategies. In this method, one first materializes the view *R_{extracted-by}*. If it is non-empty, then the *R_{like}-piece* view is materialized, and otherwise, *anchorlike-piece* is materialized. This method does as well in a batch setting as *domainlike*. In an interactive setting, it achieves a final coverage of nearly 100% with a skip rate somewhat lower than *anchorlike*.

Ranking structures with training data

Several previous researchers have considered the problem of learning wrappers from examples (Kushmerick, Weld, & Doorenbos 1997; Hsu 1998; Muslea, Minton, & Knoblock 1998). In these systems, the user provides examples of the items that should be extracted from a sample Web page, and the system induces a general procedure for extracting data from that page. If page-specific training examples are available, they can be used by storing them in a relation R , and then applying the *R_{like}* method. This use of structure-recognition methods is quite similar to previous wrapper-learning systems; one major difference, however is that no negative examples need be provided, either explicitly or implicitly.

To evaluate this wrapper-learning technique, we ran the target wrappers on each page in order to build a list of page-specific training examples. We then fixed a number of training examples m , and for each Web page, stored m randomly chosen page-specific examples in the relation R , and applied the *R_{like}* structure-recognition method. We call this the *examplelike* method. This process was repeated 10 times for each value of m , and the results were averaged.

The results from this experiment are shown in Figure 5. Even two or three labeled examples perform

⁵It seems reasonable to assume that the user (or calling program) has general knowledge about the type of the items that will be extracted. In the experiments, the items in R were always obtained from a second Web page containing items of the same type as the page being wrapped; again, it seems plausible to assume that this sort of information will be available.

	Average # Skips	Accuracy at Rank 1	Coverage at $K = \infty$
fruitful	3.3	18.3	100.0
anchorlike	0.9	69.5	100.0
domainlike	0.4	84.0	91.5
with backoff	0.6	84.0	98.8
examplelike			
$m = 2$	0.3	77.8	99.0
$m = 3$	0.3	79.3	99.3
$m = 10$	0.3	84.2	100.0
oldpagelike			
$p = 20, c = 3$	0.3	85.0	96.1
$p = 50, c = 2$	0.3	82.9	99.5
$p = 80, c = 1$	0.3	85.4	100.0

Table 1: Summary of results

somewhat better than the *anchorlike* and *fruitful* methods, and unlike the *domainlike* method, achieve complete (or nearly complete) coverage. However, the average accuracy at rank 1 is not as high as for the *domainlike* method, unless many examples are used.

These results show an advantage to presenting the user with a ranked list of wrapper pieces, as in general, coverage is improved much more by increasing K than by increasing m . For example, if the user labels two examples, then 58.6% of the pages are wrapped correctly using the top-ranked wrapper piece alone. Providing eight more examples increases coverage of the top-ranked piece to only 63.3%; however, if the user labels no additional examples, but instead considers the top two wrapper pieces, coverage jumps to 89.4%.

Maintaining a wrapper

Because Web pages frequently change, maintaining existing wrappers is a time-consuming process. In this section, we consider the problem of updating an existing wrapper for a Web page that has changed. Here a new source of information is potentially available: one could retain, for each wrapper, the data that was extracted from the previous version of the page. If the format of the page has been changed, but not its content, then the previously-extracted data can be used as page-specific training examples for the new page format, and the *examplelike* method of the previous section can be used to derive a new wrapper. If the format and content both change, then the data extracted from the old version of the page could still be used; however, it would be only an approximation to the examples that a user would provide. Using such "approximate examples" will presumably make structure-recognition more difficult; on the other hand, there will typically be many more examples than a user would provide.

Motivated by these observations, we evaluated the *R-like* structure-recognition method when R contains a large number of entries, each of which is a *corrupted* version of a data item that should be extracted from the page. Specifically, we began with a list of all data items that are extracted by the target wrapper, and

Oklahoma	dietitians
Yukon	Yukon codpiece
Vermont	Vermont
British Columbia	British Columbia Talmudizations
Oklahoma	Oklahoma
Wisconsin	Wisconsin
New Jersey	New Jersey incorrigible blubber
Alaska	Alaska
New Brunswick	
New Mexico	New Mexico cryptogram

Table 2: Ten US States and Canadian Provinces, before and after corruption with $c = 1$.

then corrupted this list as follows. First, we discarded all but randomly-chosen percentage p of the items.⁶ We next perform $c \cdot n$ random edit operations, where n is the number of retained examples. Each edit operation randomly selects one of the n items, and then either deletes a randomly chosen word from the item; or else adds a word chosen uniformly at random from */usr/dict/words*.

Figure 6 shows the results of performing this experiment (again averaged over 10 runs) with values of p ranging from 80% to 20%, and values of c ranging from 1 to 3. We call this structure-recognition method the *oldpagelike* method. With moderately corrupted example sets, the method performs very well: even with a corruption level of $p = 50\%$ and $c = 2$ it performs better on average than the *anchorlike* method.

It must be noted, however, that the corrupted examples used in this experiment are not very representative of the way a real Web page would be changed. As an illustration, Table 2 shows one list of familiar items before and after corruption with $c = 1$ (really!). It remains to be seen if more typical modifications are harder or easier to recover from.

Conclusions

In this paper, we considered the problem of recognizing "structure" in HTML pages. As formulated here, structure recognition is closely related to the task of automatically constructing wrappers: in our experiments, a "structure" is equated with a component of a wrapper, and a recognized structure is considered "meaningful" if it is part of an existing wrapper for that page. We used WHIRL, a "soft" logic that incorporates a notion of textual similarity developed in the information retrieval community, to implement several heuristic methods for recognizing structures from a narrow but useful class. Implementing these methods also required an extension to WHIRL—a "soft" version of bounded universal quantification.

Experimentally, we showed that one proposed structure-recognition method, the *anchorlike* method, performs quite well: the top-ranked structure is meaningful about 70% of the time, substantially improving

⁶Note that typically, Web pages change by having new items added, and we are trying to simulate text that would have been extracted from an old version of the page.

on simpler ranking schemes for structures, and also improving on an earlier result of ours which used more conventional methods for recognizing structure. This method is completely general, and requires no page-specific information. A second structure-recognition method, the *R-like* method, was also described, which can make use of information of many different kinds: examples of correctly-extracted text; an out-of-date version of the wrapper, together with a cached version of the last Web page that this out-of-date version correctly wrapped; or a list of objects of the same type as those that will be extracted from the Web page. In each of these cases, performance can be improved beyond that obtained by the *anchorlike* method.

References

- Ashish, N., and Knoblock, C. 1997. Wrapper generation for semistructured Internet sources. In Suciu, D., ed., *Proceedings of the Workshop on Management of Semistructured Data*. Tucson, Arizona: Available on-line from <http://www.research.att.com/~suciu/workshop-papers.html>.
- Brieman, L.; Friedman, J. H.; Olshen, R.; and Stone, C. J. 1984. *Classification and Regression Trees*. Belmont, CA: Wadsworth.
- Cohen, W. W., and Fan, W. 1999. Learning page-independent heuristics for extracting data from web pages. In *Proceedings of the 1998 AAAI Spring Symposium on Intelligent Agents in Cyberspace*.
- Cohen, W. W. 1995. Fast effective rule induction. In *Machine Learning: Proceedings of the Twelfth International Conference*. Lake Tahoe, California: Morgan Kaufmann.
- Cohen, W. W. 1998a. Integration of heterogeneous databases without common domains using queries based on textual similarity. In *Proceedings of ACM SIGMOD-98*.
- Cohen, W. W. 1998b. A Web-based information system that reasons with structured collections of text. In *Proceedings of Autonomous Agents-98*.
- Garcia-Molina, H.; Papakonstantinou, Y.; Quass, D.; Rajaraman, A.; Sagiv, Y.; Ullman, J.; and Widom, J. 1995. The TSIMMIS approach to mediation: Data models and languages (extended abstract). In *Next Generation Information Technologies and Systems (NGITS-95)*.
- Genesereth, M.; Keller, A.; and Dushka, O. 1997. Infomaster: an information integration system. In *Proceedings of the 1997 ACM SIGMOD*.
- Hammer, J.; Garcia-Molina, H.; Cho, J.; and Crespo, A. 1997. Extracting semistructured information from the Web. In Suciu, D., ed., *Proceedings of the Workshop on Management of Semistructured Data*. Tucson, Arizona: Available on-line from <http://www.research.att.com/~suciu/workshop-papers.html>.
- Hsu, C.-N. 1998. Initial results on wrapping semistructured web pages with finite-state transducers and contextual rules. In *Papers from the 1998 Workshop on AI and Information Integration*. Madison, WI: AAAI Press.
- Knoblock, C. A.; Minton, S.; Ambite, J. L.; Ashish, N.; Modi, P. J.; Muslea, I.; Philpot, A. G.; and Tejada, S. 1998. Modeling web sources for information integration. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI-98)*.
- Kushmerick, N.; Weld, D. S.; and Doorenbos, R. 1997. Wrapper induction for information extraction. In *Proceedings of the 15th International Joint Conference on Artificial Intelligence*.
- Lacroix, Z.; Sahuguet, A.; and Chandrasekar, R. 1998. User-oriented smart-cache for the web: what you seek is what you get. In *Proceedings of the 1998 ACM SIGMOD*.
- Levy, A. Y.; Rajaraman, A.; and Ordille, J. J. 1996. Querying heterogeneous information sources using source descriptions. In *Proceedings of the 22nd International Conference on Very Large Databases (VLDB-96)*.
- Mecca, G.; Atzeni, P.; Masci, A.; Merlaldo, P.; and Sindoni, G. 1998. The ARANEUS web-base management system. In *Proceedings of the 1998 ACM SIGMOD*.
- Muslea, I.; Minton, S.; and Knoblock, C. 1998. Wrapper induction for semistructured, web-based information sources. In *Proceedings of the Conference on Automated Learning and Discovery (CONALD)*.
- Porter, M. F. 1980. An algorithm for suffix stripping. *Program* 14(3):130-137.
- Salton, G., ed. 1989. *Automatic Text Processing*. Reading, Massachusetts: Addison Welsley.
- Tomasic, A.; Amouroux, R.; Bonnet, P.; and Kapitalka, O. 1997. The distributed information search component (Disco) and the World Wide Web. In *Proceedings of the 1997 ACM SIGMOD*.

This paper has been published by the American Association for Artificial Intelligence (<http://www.aaai.org>). This document has been copyrighted by the American Association for Artificial Intelligence and is being made available solely for the use and benefit of authors preparing papers for AAAI Press publications. No other use of this document may be made without permission in writing from the copyright holder.

WEST

Help

Logout

Interrupt

Main Menu

Search Form

Posting Counts

Show S Numbers

Edit S Numbers

Preferences

Cases

Search Results -

Terms	Documents
L12 and (bayesian or neural adj network)	70

Database:

US Patents Full-Text Database

US Pre-Grant Publication Full-Text Database

JPO Abstracts Database

EPO Abstracts Database

Derwent World Patents Index

IBM Technical Disclosure Bulletins

Search:

L13

Recall Text

Clear

Refine Search

Search History

DATE: Tuesday, November 04, 2003 [Printable Copy](#) [Create Case](#)

Set Name Query
side by side

Hit Count Set Name
result set

DB=USPT,PGPB,JPAB,EPAB,DWPI,TDBD; PLUR=NO; OP=OR

<u>L13</u>	L12 and (bayesian or neural adj network)	70	<u>L13</u>
<u>L12</u>	classif\$6 and structure and file and train\$6 and pattern and induct\$5 and (internet or "world wide web")	200	<u>L12</u>
<u>L11</u>	L10 and induct\$5	11	<u>L11</u>
<u>L10</u>	L9 and (internet or "world wide web")	93	<u>L10</u>
<u>L9</u>	L8 and classif\$6 and structure and file and train\$6 and pattern	124	<u>L9</u>
<u>L8</u>	((707/1 707/2 707/3 707/4 707/5 707/6 707/7 707/8 707/9 707/10)!.CCLS.)	11148	<u>L8</u>
<u>L7</u>	L1 and L2 and classif\$6 and structure and file and train\$6 and pattern	8	<u>L7</u>
<u>L6</u>	L5 and structure	62	<u>L6</u>
<u>L5</u>	L1 and classif\$6	86	<u>L5</u>
<u>L4</u>	L1 and classif\$6 and structure and file train\$6 and pattern	54231	<u>L4</u>
<u>L3</u>	L1 and L2 and classif\$6 and structure and file train\$6 and pattern	54219	<u>L3</u>
<u>L2</u>	((706/15 706/16 706/17 706/18 706/19 706/20 706/21 706/22 706/23 706/24 706/25 706/26 706/27 706/28 706/29 706/30 706/31 706/32 706/33 706/34 706/35 706/36 706/37 706/38 706/39 706/40 706/41 706/42 706/43 706/44)!.CCLS.)	2289	<u>L2</u>
<u>L1</u>	((706/12)!.CCLS.)	171	<u>L1</u>

END OF SEARCH HISTORY

WEST[Help](#)[Logout](#)[Interrupt](#)[Main Menu](#)[Search Form](#)[Posting Counts](#)[Show S Numbers](#)[Edit S Numbers](#)[Preferences](#)[Cases](#)**Search Results -**

Terms	Documents
L1 and L2 and classif\$6 and structure and file and train\$6 and pattern	8

Database:

US Patents Full-Text Database
 US Pre-Grant Publication Full-Text Database
 JPO Abstracts Database
 EPO Abstracts Database
 Derwent World Patents Index
 IBM Technical Disclosure Bulletins

Search:

L7

[Refine Search](#)[Recall Text](#)[Clear](#)**Search History**DATE: Tuesday, November 04, 2003 [Printable Copy](#) [Create Case](#)**Set Name Query**

side by side

Hit Count Set Name

result set

DB=USPT,PGPB,JPAB,EPAB,DWPI,TDBD; PLUR=NO; OP=OR

<u>L7</u>	L1 and L2 and classif\$6 and structure and file and train\$6 and pattern	8	<u>L7</u>
<u>L6</u>	L5 and structure	62	<u>L6</u>
<u>L5</u>	L1 and classif\$6	86	<u>L5</u>
<u>L4</u>	L1 and classif\$6 and structure and file train\$6 and pattern	54231	<u>L4</u>
<u>L3</u>	L1 and L2 and classif\$6 and structure and file train\$6 and pattern	54219	<u>L3</u>
<u>L2</u>	((706/15 706/16 706/17 706/18 706/19 706/20 706/21 706/22 706/23 706/24 706/25 706/26 706/27 706/28 706/29 706/30 706/31 706/32 706/33 706/34 706/35 706/36 706/37 706/38 706/39 706/40 706/41 706/42 706/43 706/44)!.CCLS.)	2289	<u>L2</u>
<u>L1</u>	((706/12)!.CCLS.)	171	<u>L1</u>

END OF SEARCH HISTORY

WEST[Generate Collection](#)[Print](#)**Search Results - Record(s) 1 through 8 of 8 returned.**☐ 1. Document ID: US 6546378 B1

L7: Entry 1 of 8

File: USPT

Apr 8, 2003

US-PAT-NO: 6546378

DOCUMENT-IDENTIFIER: US 6546378 B1

TITLE: Signal interpretation engine

DATE-ISSUED: April 8, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Cook; Daniel Reed	Bountiful	UT		

US-CL-CURRENT: 706/12; 382/159, 706/20

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments	Claims	KWC	Draw Desc	Image
----------------------	-----------------------	--------------------------	-----------------------	------------------------	--------------------------------	----------------------	---------------------------	---------------------------	-----------------------------	------------------------	---------------------	---------------------------	-----------------------

☐ 2. Document ID: US 6336109 B1

L7: Entry 2 of 8

File: USPT

Jan 1, 2002

US-PAT-NO: 6336109

DOCUMENT-IDENTIFIER: US 6336109 B1

TITLE: Method and apparatus for inducing rules from data classifiers

DATE-ISSUED: January 1, 2002

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Howard; Gary	Sawbridgeworth			GB

US-CL-CURRENT: 706/25; 706/12, 706/13, 706/2

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments	Claims	KWC	Draw Desc	Image
----------------------	-----------------------	--------------------------	-----------------------	------------------------	--------------------------------	----------------------	---------------------------	---------------------------	-----------------------------	------------------------	---------------------	---------------------------	-----------------------

☐ 3. Document ID: US 6327581 B1

L7: Entry 3 of 8

File: USPT

Dec 4, 2001

US-PAT-NO: 6327581

DOCUMENT-IDENTIFIER: US 6327581 B1

TITLE: Methods and apparatus for building a support vector machine classifier

DATE-ISSUED: December 4, 2001

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Platt; John Carlton	Bellevue	WA		

US-CL-CURRENT: 706/12; 706/14, 706/20, 706/62

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 4. Document ID: US 5970482 A

L7: Entry 4 of 8

File: USPT

Oct 19, 1999

US-PAT-NO: 5970482

DOCUMENT-IDENTIFIER: US 5970482 A

TITLE: System for data mining using neuroagents

DATE-ISSUED: October 19, 1999

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Pham; Khai Minh	Menlo Park	CA		
Rajkovic; Eric Bertrand	Foster City	CA		
Piffero; Veronique	Menlo Park	CA		

US-CL-CURRENT: 706/16; 706/12, 706/45

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 5. Document ID: US 5943662 A

L7: Entry 5 of 8

File: USPT

Aug 24, 1999

US-PAT-NO: 5943662

DOCUMENT-IDENTIFIER: US 5943662 A

TITLE: Supporting method and system for process operation

DATE-ISSUED: August 24, 1999

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Baba; Kenji	Hitachi			JP
Enbutsu; Ichiro	Hitachi			JP
Watanabe; Shoji	Hitachi			JP
Yahagi; Hayao	Hitachi			JP
Maruhashi; Fumio	Hitachi			JP
Matsuzaki; Harumi	Hitachi			JP
Matsumoto; Hiroshi	Naka-gun			JP
Nogita; Shunsuke	Hitachi			JP
Yoda; Mikio	Naka-gun			JP
Hara; Naoki	Hitachi			JP

US-CL-CURRENT: 706/23; 706/12, 706/2, 706/25

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 6. Document ID: US 5675711 A

L7: Entry 6 of 8

File: USPT

Oct 7, 1997

US-PAT-NO: 5675711

DOCUMENT-IDENTIFIER: US 5675711 A

TITLE: Adaptive statistical regression and classification of data strings, with application to the generic detection of computer viruses

DATE-ISSUED: October 7, 1997

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Kephart; Jeffrey Owen	Yorktown Heights	NY		
Sorkin; Gregory Bret	New York	NY		
Tesauro; Gerald James	Croton-on-Hudson	NY		
White; Steven Richard	New York	NY		

US-CL-CURRENT: 706/12; 706/20

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 7. Document ID: US 5438644 A

L7: Entry 7 of 8

File: USPT

Aug 1, 1995

US-PAT-NO: 5438644

DOCUMENT-IDENTIFIER: US 5438644 A

TITLE: Translation of a neural network into a rule-based expert system

DATE-ISSUED: August 1, 1995

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Fu; Li M.	Gainesville	FL		

US-CL-CURRENT: 706/12; 706/25, 706/45

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 8. Document ID: US 5396580 A

L7: Entry 8 of 8

File: USPT

Mar 7, 1995

US-PAT-NO: 5396580

DOCUMENT-IDENTIFIER: US 5396580 A

TITLE: Translation of a neural network into a rule-based expert system

DATE-ISSUED: March 7, 1995

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Fu; Li M.	Gainesville	FL		

US-CL-CURRENT: 706/12; 706/25, 706/45

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

Generate Collection

Print

Terms	Documents
L1 and L2 and classif\$6 and structure and file and train\$6 and pattern	8

Display Format:

-

Change Format

Previous Page

Next Page

WEST[Generate Collection](#)[Print](#)**Search Results - Record(s) 1 through 8 of 8 returned.**☐ 1. Document ID: US 6546378 B1

L7: Entry 1 of 8

File: USPT

Apr 8, 2003

US-PAT-NO: 6546378

DOCUMENT-IDENTIFIER: US 6546378 B1

TITLE: Signal interpretation engine

DATE-ISSUED: April 8, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Cook; Daniel Reed	Bountiful	UT		

US-CL-CURRENT: 706/12; 382/159, 706/20[Full](#) | [Title](#) | [Citation](#) | [Front](#) | [Review](#) | [Classification](#) | [Date](#) | [Reference](#) | [Sequences](#) | [Attachments](#)[KIMC](#) | [Draw Desc](#) | [Image](#)☒ 2. Document ID: US 6336109 B1

L7: Entry 2 of 8

File: USPT

Jan 1, 2002

US-PAT-NO: 6336109

DOCUMENT-IDENTIFIER: US 6336109 B1

TITLE: Method and apparatus for inducing rules from data classifiers

DATE-ISSUED: January 1, 2002

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Howard; Gary	Sawbridgeworth			GB

US-CL-CURRENT: 706/25; 706/12, 706/13, 706/2[Full](#) | [Title](#) | [Citation](#) | [Front](#) | [Review](#) | [Classification](#) | [Date](#) | [Reference](#) | [Sequences](#) | [Attachments](#)[KIMC](#) | [Draw Desc](#) | [Image](#)☒ 3. Document ID: US 6327581 B1

L7: Entry 3 of 8

File: USPT

Dec 4, 2001

US-PAT-NO: 6327581

DOCUMENT-IDENTIFIER: US 6327581 B1

TITLE: Methods and apparatus for building a support vector machine classifier

DATE-ISSUED: December 4, 2001

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Platt; John Carlton	Bellevue	WA		

US-CL-CURRENT: 706/12; 706/14, 706/20, 706/62

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☒ 4. Document ID: US 5970482 A

L7: Entry 4 of 8

File: USPT

Oct 19, 1999

US-PAT-NO: 5970482

DOCUMENT-IDENTIFIER: US 5970482 A

TITLE: System for data mining using neuroagents

DATE-ISSUED: October 19, 1999

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Pham; Khai Minh	Menlo Park	CA		
Rajkovic; Eric Bertrand	Foster City	CA		
Piffero; Veronique	Menlo Park	CA		

US-CL-CURRENT: 706/16; 706/12, 706/45

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☒ 5. Document ID: US 5943662 A

L7: Entry 5 of 8

File: USPT

Aug 24, 1999

US-PAT-NO: 5943662

DOCUMENT-IDENTIFIER: US 5943662 A

TITLE: Supporting method and system for process operation

DATE-ISSUED: August 24, 1999

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Baba; Kenji	Hitachi			JP
Enbutsu; Ichiro	Hitachi			JP
Watanabe; Shoji	Hitachi			JP
Yahagi; Hayao	Hitachi			JP
Maruhashi; Fumio	Hitachi			JP
Matsuzaki; Harumi	Hitachi			JP
Matsumoto; Hiroshi	Naka-gun			JP
Nogita; Shunsuke	Hitachi			JP
Yoda; Mikio	Naka-gun			JP
Hara; Naoki	Hitachi			JP

US-CL-CURRENT: 706/23; 706/12, 706/2, 706/25

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KMIC	Draw Desc	Image
------	-----------	-------

☒ 6. Document ID: US 5675711 A

L7: Entry 6 of 8

File: USPT

Oct 7, 1997

US-PAT-NO: 5675711

DOCUMENT-IDENTIFIER: US 5675711 A

TITLE: Adaptive statistical regression and classification of data strings, with application to the generic detection of computer viruses

DATE-ISSUED: October 7, 1997

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Kephart; Jeffrey Owen	Yorktown Heights	NY		
Sorkin; Gregory Bret	New York	NY		
Tesauro; Gerald James	Croton-on-Hudson	NY		
White; Steven Richard	New York	NY		

US-CL-CURRENT: 706/12; 706/20

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KMIC	Draw Desc	Image
------	-----------	-------

☐ 7. Document ID: US 5438644 A

L7: Entry 7 of 8

File: USPT

Aug 1, 1995

US-PAT-NO: 5438644

DOCUMENT-IDENTIFIER: US 5438644 A

TITLE: Translation of a neural network into a rule-based expert system

DATE-ISSUED: August 1, 1995

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Fu; Li M.	Gainesville	FL		

US-CL-CURRENT: 706/12; 706/25, 706/45

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KMIC	Draw Desc	Image
------	-----------	-------

☐ 8. Document ID: US 5396580 A

L7: Entry 8 of 8

File: USPT

Mar 7, 1995

US-PAT-NO: 5396580

DOCUMENT-IDENTIFIER: US 5396580 A

TITLE: Translation of a neural network into a rule-based expert system

DATE-ISSUED: March 7, 1995

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Fu; Li M.	Gainesville	FL		

US-CL-CURRENT: 706/12; 706/25, 706/45

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

RMIC	Draw Desc	Image
------	-----------	-------

Generate Collection

Print

Terms	Documents
L1 and L2 and classif\$6 and structure and file and train\$6 and pattern	8

Display Format:

-

Change Format

Previous Page

Next Page

WEST[Help](#)[Logout](#)[Interrupt](#)[Main Menu](#)[Search Form](#)[Posting Counts](#)[Show S Numbers](#)[Edit S Numbers](#)[Preferences](#)[Cases](#)**Search Results -**

Terms	Documents
L1 and L2 and classif\$6 and structure and file and train\$6 and pattern	8

Database:

US Patents Full-Text Database
 US Pre-Grant Publication Full-Text Database
 JPO Abstracts Database
 EPO Abstracts Database
 Derwent World Patents Index
 IBM Technical Disclosure Bulletins

Search:

L7

[Refine Search](#)[Recall Text](#)[Clear](#)**Search History**
DATE: Tuesday, November 04, 2003
[Printable Copy](#)
[Create Case](#)
Set Name Query
 side by side

Hit Count Set Name
 result set

DB=USPT,PGPB,JPAB,EPAB,DWPI,TDBD; PLUR=NO; OP=OR

<u>L7</u>	L1 and L2 and classif\$6 and structure and file and train\$6 and pattern	8	<u>L7</u>
<u>L6</u>	L5 and structure	62	<u>L6</u>
<u>L5</u>	L1 and classif\$6	86	<u>L5</u>
<u>L4</u>	L1 and classif\$6 and structure and file train\$6 and pattern	54231	<u>L4</u>
<u>L3</u>	L1 and L2 and classif\$6 and structure and file train\$6 and pattern	54219	<u>L3</u>
<u>L2</u>	((706/15 706/16 706/17 706/18 706/19 706/20 706/21 706/22 706/23 706/24 706/25 706/26 706/27 706/28 706/29 706/30 706/31 706/32 706/33 706/34 706/35 706/36 706/37 706/38 706/39 706/40 706/41 706/42 706/43 706/44)!.CCLS.)	2289	<u>L2</u>
<u>L1</u>	((706/12)!.CCLS.)	171	<u>L1</u>

END OF SEARCH HISTORY

 [Home](#) | [About](#) | [Privacy](#) | [Contact](#) | [Feedback](#)

[Patent Intranet](#) > [Classification Home Page](#) > [Classification Search Page](#) > [Site Feedback](#)

[Classification Schedule](#)

[Search Classification Data](#) | [Class Numbers & Titles](#) | [Class Numbers](#) | [USPC Index](#) | [International](#) | [HELP](#) | [Employee by Name](#) | [Employees by Org](#)

[<-Previous Page](#)

Class 707 DATA PROCESSING: DATABASE AND FILE MANAGEMENT OR DATA STRUCTURES

[Click here to view a PDF version of this file](#)

- 1** **DATABASE OR FILE ACCESSING**
- 2 . Access augmentation or optimizing
- 3 . Query processing (i.e., searching)
- 4 .. Query formulation, input preparation, or translation
- 5 .. Query augmenting and refining (e.g., inexact access)
- 6 .. Pattern matching access
- 7 . Sorting
- 8 . Concurrency (e.g., lock management in shared database)
- 9 . Privileged access
- 10 . Distributed or remote access
- 100** **DATABASE SCHEMA OR DATA STRUCTURE**
- 101 . Manipulating data structure (e.g., compression, compaction, compilation)
- 102 . Generating database or data structure (e.g., via user interface)
- 103R . Object-oriented database structure
- 103Y .. Object-oriented database structure processing
- 103X .. Object-oriented database structure network
- 103Z .. Object-oriented database structure reference
- 104.1 . Application of database or data structure (e.g., distributed, multimedia, image)

- 200** **FILE OR DATABASE MAINTENANCE**
- 201 . Coherency (e.g., same view to multiple users)
- 202 .. Recoverability
- 203 .. Version management
- 204 .. Archiving or backup
- 205 . File allocation
- 206 .. Garbage collection

FOREIGN ART COLLECTIONS

FOR000 CLASS-RELATED FOREIGN DOCUMENTS

Note: Some content linked to on this page may require a plug-in for Adobe Acrobat Reader.

This file produced by USPTO - SIRA - Office of Patent Automation - ReferenceTools Project. Questions or comments relating to this file should be directed to [Patent Automation Feedback](#).

[Intranet Home](#) | [Index](#) | [Resources](#) | [Contacts](#) | [Internet](#) | [Search](#) | [Firewall](#) | [Web Services](#)

Last Modified: Tuesday, October 07, 2003 12:13:04

This data is current as of August/2003

WEST[Help](#)[Logout](#)[Interrupt](#)[Main Menu](#)[Search Form](#)[Posting Counts](#)[Show S Numbers](#)[Edit S Numbers](#)[Preferences](#)[Cases](#)**Search Results -**

Terms	Documents
L9 and (internet or "world wide web")	93

Database:

US Patents Full-Text Database
US Pre-Grant Publication Full-Text Database
JPO Abstracts Database
EPO Abstracts Database
Derwent World Patents Index
IBM Technical Disclosure Bulletins

Search:

L10

[Refine Search](#)[Recall Text](#)[Clear](#)**Search History****DATE:** Tuesday, November 04, 2003 [Printable Copy](#) [Create Case](#)

Set Name Query
side by side

Hit Count Set Name
result set

DB=USPT,PGPB,JPAB,EPAB,DWPI,TDBD; PLUR=NO; OP=OR

<u>L10</u>	L9 and (internet or "world wide web")	93	<u>L10</u>
<u>L9</u>	L8 and classif\$6 and structure and file and train\$6 and pattern	124	<u>L9</u>
<u>L8</u>	((707/1 707/2 707/3 707/4 707/5 707/6 707/7 707/8 707/9 707/10)!.CCLS.)	11148	<u>L8</u>
<u>L7</u>	L1 and L2 and classif\$6 and structure and file and train\$6 and pattern	8	<u>L7</u>
<u>L6</u>	L5 and structure	62	<u>L6</u>
<u>L5</u>	L1 and classif\$6	86	<u>L5</u>
<u>L4</u>	L1 and classif\$6 and structure and file train\$6 and pattern	54231	<u>L4</u>
<u>L3</u>	L1 and L2 and classif\$6 and structure and file train\$6 and pattern	54219	<u>L3</u>
<u>L2</u>	((706/15 706/16 706/17 706/18 706/19 706/20 706/21 706/22 706/23 706/24 706/25 706/26 706/27 706/28 706/29 706/30 706/31 706/32 706/33 706/34 706/35 706/36 706/37 706/38 706/39 706/40 706/41 706/42 706/43 706/44)!.CCLS.)	2289	<u>L2</u>
<u>L1</u>	((706/12)!.CCLS.)	171	<u>L1</u>

END OF SEARCH HISTORY

WEST[Help](#)[Logout](#)[Interrupt](#)[Main Menu](#)[Search Form](#)[Posting Counts](#)[Show S Numbers](#)[Edit S Numbers](#)[Preferences](#)[Cases](#)**Search Results -**

Terms	Documents
L10 and induct\$5	11

Database:

US Patents Full-Text Database
US Pre-Grant Publication Full-Text Database
JPO Abstracts Database
EPO Abstracts Database
Derwent World Patents Index
IBM Technical Disclosure Bulletins

Search:

L11

[Refine Search](#)[Recall Text](#)[Clear](#)**Search History****DATE:** Tuesday, November 04, 2003 [Printable Copy](#) [Create Case](#)

Set Name Query
side by side

Hit Count Set Name
result set

DB=USPT,PGPB,JPAB,EPAB,DWPI,TDBD; PLUR=NO; OP=OR

<u>L11</u>	L10 and induct\$5	11	<u>L11</u>
<u>L10</u>	L9 and (internet or "world wide web")	93	<u>L10</u>
<u>L9</u>	L8 and classif\$6 and structure and file and train\$6 and pattern	124	<u>L9</u>
<u>L8</u>	((707/1 707/2 707/3 707/4 707/5 707/6 707/7 707/8 707/9 707/10)!.CCLS.)	11148	<u>L8</u>
<u>L7</u>	L1 and L2 and classif\$6 and structure and file and train\$6 and pattern	8	<u>L7</u>
<u>L6</u>	L5 and structure	62	<u>L6</u>
<u>L5</u>	L1 and classif\$6	86	<u>L5</u>
<u>L4</u>	L1 and classif\$6 and structure and file train\$6 and pattern	54231	<u>L4</u>
<u>L3</u>	L1 and L2 and classif\$6 and structure and file train\$6 and pattern	54219	<u>L3</u>
<u>L2</u>	((706/15 706/16 706/17 706/18 706/19 706/20 706/21 706/22 706/23 706/24 706/25 706/26 706/27 706/28 706/29 706/30 706/31 706/32 706/33 706/34 706/35 706/36 706/37 706/38 706/39 706/40 706/41 706/42 706/43 706/44)!.CCLS.)	2289	<u>L2</u>
<u>L1</u>	((706/12)!.CCLS.)	171	<u>L1</u>

END OF SEARCH HISTORY

WEST[Generate Collection](#)[Print](#)**Search Results - Record(s) 1 through 11 of 11 returned.**☐ 1. Document ID: US 20030167209 A1

L11: Entry 1 of 11

File: PGPB

Sep 4, 2003

PGPUB-DOCUMENT-NUMBER: 20030167209

PGPUB-FILING-TYPE: new

DOCUMENT-IDENTIFIER: US 20030167209 A1

TITLE: Online intelligent information comparison agent of multilingual electronic data sources over inter-connected computer networks

PUBLICATION-DATE: September 4, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Hsieh, Victor	Alhambra	CA	US	

US-CL-CURRENT: 705/26; 705/28, 707/10, 707/9[Full](#) | [Title](#) | [Citation](#) | [Front](#) | [Review](#) | [Classification](#) | [Date](#) | [Reference](#) | [Sequences](#) | [Attachments](#)[KIMC](#) | [Draw Desc](#) | [Image](#)☐ 2. Document ID: US 20030069877 A1

L11: Entry 2 of 11

File: PGPB

Apr 10, 2003

PGPUB-DOCUMENT-NUMBER: 20030069877

PGPUB-FILING-TYPE: new

DOCUMENT-IDENTIFIER: US 20030069877 A1

TITLE: System for automatically generating queries

PUBLICATION-DATE: April 10, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Grefenstette, Gregory T.	Gieres	PA	FR	
Shanahan, James G.	Pittsburgh		US	

US-CL-CURRENT: 707/2[Full](#) | [Title](#) | [Citation](#) | [Front](#) | [Review](#) | [Classification](#) | [Date](#) | [Reference](#) | [Sequences](#) | [Attachments](#)[KIMC](#) | [Draw Desc](#) | [Image](#)☐ 3. Document ID: US 20030061201 A1

L11: Entry 3 of 11

File: PGPB

Mar 27, 2003

PGPUB-DOCUMENT-NUMBER: 20030061201

PGPUB-FILING-TYPE: new
DOCUMENT-IDENTIFIER: US 20030061201 A1

TITLE: System for propagating enrichment between documents

PUBLICATION-DATE: March 27, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Grefenstette, Gregory T.	Gieres	PA	FR	
Shanahan, James G.	Pittsburgh		US	

US-CL-CURRENT: 707/3

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 4. Document ID: US 20030061200 A1

L11: Entry 4 of 11

File: PGPB

Mar 27, 2003

PGPUB-DOCUMENT-NUMBER: 20030061200
PGPUB-FILING-TYPE: new
DOCUMENT-IDENTIFIER: US 20030061200 A1

TITLE: System with user directed enrichment and import/export control

PUBLICATION-DATE: March 27, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Hubert, Laurence	St Bernard du Touvet		FR	
Guerin, Nicolas	Grenoble		FR	

US-CL-CURRENT: 707/3

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 5. Document ID: US 20030033288 A1

L11: Entry 5 of 11

File: PGPB

Feb 13, 2003

PGPUB-DOCUMENT-NUMBER: 20030033288
PGPUB-FILING-TYPE: new
DOCUMENT-IDENTIFIER: US 20030033288 A1

TITLE: Document-centric system with auto-completion and auto-correction

PUBLICATION-DATE: February 13, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Shanahan, James G.	Pittsburgh	PA	US	
Grefenstette, Gregory T.	Gieres		FR	

US-CL-CURRENT: 707/3

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 6. Document ID: US 20030033287 A1

L11: Entry 6 of 11

File: PGPB

Feb 13, 2003

PGPUB-DOCUMENT-NUMBER: 20030033287

PGPUB-FILING-TYPE: new

DOCUMENT-IDENTIFIER: US 20030033287 A1

TITLE: Meta-document management system with user definable personalities

PUBLICATION-DATE: February 13, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Shanahan, James G.	Pittsburgh	PA	US	
Grefenstette, Gregory T.	Gieres		FR	
Fernstrom, Christer	St-Ismier		FR	
Hubert, Laurence	St Bernard du Touvet		FR	

US-CL-CURRENT: 707/3

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 7. Document ID: US 20020107843 A1

L11: Entry 7 of 11

File: PGPB

Aug 8, 2002

PGPUB-DOCUMENT-NUMBER: 20020107843

PGPUB-FILING-TYPE: new

DOCUMENT-IDENTIFIER: US 20020107843 A1

TITLE: Customer self service subsystem for classifying user contexts

PUBLICATION-DATE: August 8, 2002

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Biebesheimer, Debra L.	Carmel	NY	US	
Keller, Neal M.	Somers	NY	US	
Oblinger, Daniel A.	New York	NY	US	
Podlaseck, Mark E.	New Preston	CT	US	
Rolando, Stephen J.	Katonah	NY	US	

US-CL-CURRENT: 707/3

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 8. Document ID: US 20010049671 A1

L11: Entry 8 of 11

File: PGPB

Dec 6, 2001

PGPUB-DOCUMENT-NUMBER: 20010049671
PGPUB-FILING-TYPE: new
DOCUMENT-IDENTIFIER: US 20010049671 A1

TITLE: e-Stract: a process for knowledge-based retrieval of electronic information

PUBLICATION-DATE: December 6, 2001

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Joerg, Werner B.	Salt Lake City	UT	US	

US-CL-CURRENT: 706/50; 707/3

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 9. Document ID: US 6230312 B1

L11: Entry 9 of 11

File: USPT

May 8, 2001

US-PAT-NO: 6230312
DOCUMENT-IDENTIFIER: US 6230312 B1
**** See image for Certificate of Correction ****

TITLE: Automatic detection of per-unit location constraints

DATE-ISSUED: May 8, 2001

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Hunt; Galen C.	Bellevue	WA		

US-CL-CURRENT: 717/108; 707/10, 707/104.1, 709/312, 709/331, 717/142, 717/163

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 10. Document ID: US 6026388 A

L11: Entry 10 of 11

File: USPT

Feb 15, 2000

US-PAT-NO: 6026388
DOCUMENT-IDENTIFIER: US 6026388 A

TITLE: User interface and other enhancements for natural language information retrieval system and method

DATE-ISSUED: February 15, 2000

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Liddy; Elizabeth D.	Syracuse	NY		
Paik; Woojin	Syracuse	NY		
McKenna; Mary E.	Syracuse	NY		
Weiner; Michael L.	Webster	NY		
Yu; Edmund S.	DeWitt	NY		
Diamond; Theodore G.	Mercer Island	WA		
Balakrishnan; Bhaskaran	Syracuse	NY		
Snyder; David L.	Pittsford	NY		

US-CL-CURRENT: 707/1; 704/9, 707/5

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWC	Draw Desc	Image
-----	-----------	-------

☐ 11. Document ID: US 5963940 A

L11: Entry 11 of 11

File: USPT

Oct 5, 1999

US-PAT-NO: 5963940

DOCUMENT-IDENTIFIER: US 5963940 A

TITLE: Natural language information retrieval system and method

DATE-ISSUED: October 5, 1999

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Liddy; Elizabeth D.	Syracuse	NY		
Paik; Woojin	Syracuse	NY		
McKenna; Mary E.	Syracuse	NY		
Li; Ming	Jersey City	NJ		

US-CL-CURRENT: 707/5; 704/9

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWC	Draw Desc	Image
-----	-----------	-------

Terms	Documents
L10 and induct\$5	11

Display Format: [Previous Page](#)[Next Page](#)

WEST[Help](#)[Logout](#)[Interrupt](#)[Main Menu](#)[Search Form](#)[Posting Counts](#)[Show S Numbers](#)[Edit S Numbers](#)[Preferences](#)[Cases](#)**Search Results -**

Terms	Documents
L10 and induct\$5	11

Database:

US Patents Full-Text Database
US Pre-Grant Publication Full-Text Database
JPO Abstracts Database
EPO Abstracts Database
Derwent World Patents Index
IBM Technical Disclosure Bulletins

Search:

L11

[Refine Search](#)[Recall Text](#)[Clear](#)**Search History****DATE:** Tuesday, November 04, 2003 [Printable Copy](#) [Create Case](#)

Set Name Query
side by side

Hit Count Set Name
result set

DB=USPT,PGPB,JPAB,EPAB,DWPI,TDBD; PLUR=NO; OP=OR

<u>L11</u>	L10 and induct\$5	11	<u>L11</u>
<u>L10</u>	L9 and (internet or "world wide web")	93	<u>L10</u>
<u>L9</u>	L8 and classif\$6 and structure and file and train\$6 and pattern	124	<u>L9</u>
<u>L8</u>	((707/1 707/2 707/3 707/4 707/5 707/6 707/7 707/8 707/9 707/10)!.CCLS.)	11148	<u>L8</u>
<u>L7</u>	L1 and L2 and classif\$6 and structure and file and train\$6 and pattern	8	<u>L7</u>
<u>L6</u>	L5 and structure	62	<u>L6</u>
<u>L5</u>	L1 and classif\$6	86	<u>L5</u>
<u>L4</u>	L1 and classif\$6 and structure and file train\$6 and pattern	54231	<u>L4</u>
<u>L3</u>	L1 and L2 and classif\$6 and structure and file train\$6 and pattern	54219	<u>L3</u>
<u>L2</u>	((706/15 706/16 706/17 706/18 706/19 706/20 706/21 706/22 706/23 706/24 706/25 706/26 706/27 706/28 706/29 706/30 706/31 706/32 706/33 706/34 706/35 706/36 706/37 706/38 706/39 706/40 706/41 706/42 706/43 706/44)!.CCLS.)	2289	<u>L2</u>
<u>L1</u>	((706/12)!.CCLS.)	171	<u>L1</u>

END OF SEARCH HISTORY

WEST[Help](#)[Logout](#)[Interrupt](#)[Main Menu](#)[Search Form](#)[Posting Counts](#)[Show S Numbers](#)[Edit S Numbers](#)[Preferences](#)[Cases](#)**Search Results -**

Terms	Documents
L12 and (bayesian or neural adj network)	70

Database:

US Patents Full-Text Database
US Pre-Grant Publication Full-Text Database
JPO Abstracts Database
EPO Abstracts Database
Derwent World Patents Index
IBM Technical Disclosure Bulletins

Search:

L13

[Refine Search](#)[Recall Text](#)[Clear](#)**Search History****DATE:** Tuesday, November 04, 2003 [Printable Copy](#) [Create Case](#)

Set Name Query
side by side

Hit Count Set Name
result set

DB=USPT,PGPB,JPAB,EPAB,DWPI,TDBD; PLUR=NO; OP=OR

<u>L13</u>	L12 and (bayesian or neural adj network)	70	<u>L13</u>
<u>L12</u>	classif\$6 and structure and file and train\$6 and pattern and induct\$5 and (internet or "world wide web")	200	<u>L12</u>
<u>L11</u>	L10 and induct\$5	11	<u>L11</u>
<u>L10</u>	L9 and (internet or "world wide web")	93	<u>L10</u>
<u>L9</u>	L8 and classif\$6 and structure and file and train\$6 and pattern	124	<u>L9</u>
<u>L8</u>	((707/1 707/2 707/3 707/4 707/5 707/6 707/7 707/8 707/9 707/10)!.CCLS.)	11148	<u>L8</u>
<u>L7</u>	L1 and L2 and classif\$6 and structure and file and train\$6 and pattern	8	<u>L7</u>
<u>L6</u>	L5 and structure	62	<u>L6</u>
<u>L5</u>	L1 and classif\$6	86	<u>L5</u>
<u>L4</u>	L1 and classif\$6 and structure and file train\$6 and pattern	54231	<u>L4</u>
<u>L3</u>	L1 and L2 and classif\$6 and structure and file train\$6 and pattern	54219	<u>L3</u>
<u>L2</u>	((706/15 706/16 706/17 706/18 706/19 706/20 706/21 706/22 706/23 706/24 706/25 706/26 706/27 706/28 706/29 706/30 706/31 706/32 706/33 706/34 706/35 706/36 706/37 706/38 706/39 706/40 706/41 706/42 706/43 706/44)!.CCLS.)	2289	<u>L2</u>
<u>L1</u>	((706/12)!.CCLS.)	171	<u>L1</u>

END OF SEARCH HISTORY

WEST[Generate Collection](#)[Print](#)**Search Results - Record(s) 1 through 50 of 70 returned.**☐ 1. Document ID: US 20030199685 A1

L13: Entry 1 of 70

File: PGPB

Oct 23, 2003

PGPUB-DOCUMENT-NUMBER: 20030199685

PGPUB-FILING-TYPE: new

DOCUMENT-IDENTIFIER: US 20030199685 A1

TITLE: Cell-based detection and differentiation of disease states

PUBLICATION-DATE: October 23, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Pressman, Norman J.	Glencoe	IL	US	
Hirsch, Kenneth S.	Redwood City	CA	US	
Hirsch, Adrian	Redwood City	CA	US	

US-CL-CURRENT: [536/24.3](#); [435/69.1](#), [435/91.1](#)

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments	Claims	KMC	Draw Desc	Image
----------------------	-----------------------	--------------------------	-----------------------	------------------------	--------------------------------	----------------------	---------------------------	---------------------------	-----------------------------	------------------------	---------------------	---------------------------	-----------------------

☐ 2. Document ID: US 20030191608 A1

L13: Entry 2 of 70

File: PGPB

Oct 9, 2003

PGPUB-DOCUMENT-NUMBER: 20030191608

PGPUB-FILING-TYPE: new

DOCUMENT-IDENTIFIER: US 20030191608 A1

TITLE: Data processing and observation system

PUBLICATION-DATE: October 9, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Anderson, Mark Stephen	Edinburgh		AU	
Engelhardt, Dean Crawford	Edinburgh		AU	
Marriott, Damian Andrew	Edinburgh		AU	
Randhawa, Suneel Singh	Edinburgh		AU	

US-CL-CURRENT: [702/189](#)

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments	Claims	KMC	Draw Desc	Image
----------------------	-----------------------	--------------------------	-----------------------	------------------------	--------------------------------	----------------------	---------------------------	---------------------------	-----------------------------	------------------------	---------------------	---------------------------	-----------------------

☐ 3. Document ID: US 20030190602 A1

L13: Entry 3 of 70

File: PGPB

Oct 9, 2003

PGPUB-DOCUMENT-NUMBER: 20030190602
PGPUB-FILING-TYPE: new
DOCUMENT-IDENTIFIER: US 20030190602 A1

TITLE: Cell-based detection and differentiation of disease states

PUBLICATION-DATE: October 9, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Pressman, Norman J.	Glencoe	IL	US	
Hirsch, Kenneth S.	Redwood City	CA	US	

US-CL-CURRENT: 435/5; 435/287.2, 435/6, 435/7.23, 435/7.92

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 4. Document ID: US 20030187642 A1

L13: Entry 4 of 70

File: PGPB

Oct 2, 2003

PGPUB-DOCUMENT-NUMBER: 20030187642
PGPUB-FILING-TYPE: new
DOCUMENT-IDENTIFIER: US 20030187642 A1

TITLE: System and method for the automatic discovery of salient segments in speech transcripts

PUBLICATION-DATE: October 2, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Ponceleon, Dulce Beatriz	Palo Alto	CA	US	
Srinivasan, Savitha	San Jose	CA	US	

US-CL-CURRENT: 704/252

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 5. Document ID: US 20030176931 A1

L13: Entry 5 of 70

File: PGPB

Sep 18, 2003

PGPUB-DOCUMENT-NUMBER: 20030176931
PGPUB-FILING-TYPE: new
DOCUMENT-IDENTIFIER: US 20030176931 A1

TITLE: Method for constructing segmentation-based predictive models

PUBLICATION-DATE: September 18, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Pednault, Edwin Peter Dawson	Cortlandt Manor	NY	US	
Natarajan, Ramesh	Pleasantville	NY	US	

US-CL-CURRENT: 700/31; 700/30, 700/44

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

WWW	Draw Desc	Image
-----	-----------	-------

☐ 6. Document ID: US 20030158795 A1

L13: Entry 6 of 70

File: PGPB

Aug 21, 2003

PGPUB-DOCUMENT-NUMBER: 20030158795
PGPUB-FILING-TYPE: new
DOCUMENT-IDENTIFIER: US 20030158795 A1

TITLE: Quality management and intelligent manufacturing with labels and smart tags
in event-based product manufacturing

PUBLICATION-DATE: August 21, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Markham, Charles Earl	Appleton	WI	US	
Barber, Douglas Gordon Barron	Appleton	WI	US	
Hise, John Harland	Neenah	WI	US	
Ihde, Sheryl Annette	Greenville	WI	US	
Lindsay, Jeffrey Dean	Appleton	WI	US	
Nygaard, Kurt Sigurd	Appleton	WI	US	
Pokorny, Michael Roy	Neenah	WI	US	
Price, Michael T.	Oshkosh	WI	US	
Reade, Walter Caswell	Appleton	WI	US	
Shaffer, Gregory Duncan	Neenah	WI	US	
Yosten, Roger Dale	Sumner	TX	US	

US-CL-CURRENT: 705/28

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

WWW	Draw Desc	Image
-----	-----------	-------

☐ 7. Document ID: US 20030155415 A1

L13: Entry 7 of 70

File: PGPB

Aug 21, 2003

PGPUB-DOCUMENT-NUMBER: 20030155415
PGPUB-FILING-TYPE: new
DOCUMENT-IDENTIFIER: US 20030155415 A1

TITLE: Communication between machines and feed-forward control in event-based
product manufacturing

PUBLICATION-DATE: August 21, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Markham, Charles Earl	Appleton	WI	US	
Barber, Douglas Gordon Barron	Appleton	WI	US	
Fuller, Paul D.	Menasha	WI	US	
Hise, John Harland	Neenah	WI	US	
Ihde, Sheryl Annette	Greenville	WI	US	
Lindsay, Jeffrey Dean	Appleton	WI	US	
Matheus, Jon Ray	Appleton	WI	US	
Nygaard, Kurt Sigurd	Appleton	WI	US	
Pokorny, Michael Roy	Neenah	WI	US	
Reads, Walter Caswell	Appleton	WI	US	
Shaffer, Gregory Duncan	Neenah	WI	US	
Tiffany, Flynn Matthew	Layton	UT	US	
Yosten, Roger Dale	Sumner	TX	US	

US-CL-CURRENT: 235/376

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 8. Document ID: US 20030154144 A1

L13: Entry 8 of 70

File: PGPB

Aug 14, 2003

PGPUB-DOCUMENT-NUMBER: 20030154144

PGPUB-FILING-TYPE: new

DOCUMENT-IDENTIFIER: US 20030154144 A1

TITLE: Integrating event-based production information with financial and purchasing systems in product manufacturing

PUBLICATION-DATE: August 14, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Pokorny, Michael Roy	Neenah	WI	US	
Barber, Douglas Gordon Barron	Appleton	WI	US	
Bush, Perry A.	Menasha	WI	US	
Hise, John Harland	Neenah	WI	US	
Shun Hoo, Winnie Shi Mei	Appleton	WI	US	
Markham, Charles Earl	Appleton	WI	US	
Matheus, Jon Ray	Appleton	WI	US	
Mork, Jamie Scott	Greenville	WI	US	
Nygaard, Kurt Sigurd	Appleton	WI	US	
Shaffer, Gregory Duncan	Neenah	WI	US	
Stambuk, Jose Andres	Appleton	WI	US	

US-CL-CURRENT: 705/28

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 9. Document ID: US 20030150909 A1

L13: Entry 9 of 70

File: PGPB

Aug 14, 2003

PGPUB-DOCUMENT-NUMBER: 20030150909
PGPUB-FILING-TYPE: new
DOCUMENT-IDENTIFIER: US 20030150909 A1

TITLE: Quality management by validating a bill of materials in event-based product manufacturing

PUBLICATION-DATE: August 14, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Markham, Charles Earl	Appleton	WI	US	
Barber, Douglas Gordon Barron	Appleton	WI	US	
Boyd, Amy Hancock	Neenah	WI	US	
Goggans, Gary Lee	Appleton	WI	US	
Hise, John Harland	Neenah	WI	US	
Ihde, Sheryl Annette	Greenville	WI	US	
Lindsay, Jeffrey Dean	Appleton	WI	US	
Meissner, Jolene Marie	DePere	WI	US	
Mork, Jamie Scott	Greenville	WI	US	
Nygaard, Kurt Sigurd	Appleton	WI	US	
Park, Scott Jeffrey	Menasha	WI	US	
Pokorny, Michael Roy	Neenah	WI	US	
Reade, Walter Caswell	Appleton	WI	US	
Reynders, John L.	Appleton	WI	US	
Shaffer, Gregory Duncan	Neenah	WI	US	
Yosten, Roger Dale	Sumner	TX	US	

US-CL-CURRENT: 235/376

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 10. Document ID: US 20030150908 A1

L13: Entry 10 of 70

File: PGPB

Aug 14, 2003

PGPUB-DOCUMENT-NUMBER: 20030150908
PGPUB-FILING-TYPE: new
DOCUMENT-IDENTIFIER: US 20030150908 A1

TITLE: User interface for reporting event-based production information in product manufacturing

PUBLICATION-DATE: August 14, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Pokorny, Michael Roy	Neenah	WI	US	
Barber, Douglas Gordon Barron	Appleton	WI	US	
Hoo, Winnie Shi Mei Shun	Appleton	WI	US	
Lindsay, Jeffrey Dean	Appleton	WI	US	
Markham, Charles Earl	Appleton	WI	US	
Matheus, Jon Ray	Appleton	WI	US	
Mork, Jamie Scott	Greenville	WI	US	
Nygaard, Kurt Sigurd	Appleton	WI	US	
Reade, Walter Caswell	Appleton	WI	US	
Stambuk, Jose Andres	Appleton	WI	US	

US-CL-CURRENT: 235/375

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 11. Document ID: US 20030144746 A1

L13: Entry 11 of 70

File: PGPB

Jul 31, 2003

PGPUB-DOCUMENT-NUMBER: 20030144746
PGPUB-FILING-TYPE: new
DOCUMENT-IDENTIFIER: US 20030144746 A1

TITLE: Control for an industrial process using one or more multidimensional variables

PUBLICATION-DATE: July 31, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Hsiung, Chang-Meng	Irvine	CA	US	
Munoz, Bethsabeth	Pasadena	CA	US	
Roy, Ajoy	Pasadena	CA	US	
Steinthal, Michael	Los Angeles	CA	US	
Sunshine, Steven	Pasadena	CA	US	
Vicic, Michael Allen	Pasadena	CA	US	
Zhang, Shou-Hua	Arcadia	CA	US	

US-CL-CURRENT: 700/28; 700/26, 700/31, 700/32

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 12. Document ID: US 20030112921 A1

L13: Entry 12 of 70

File: PGPB

Jun 19, 2003

PGPUB-DOCUMENT-NUMBER: 20030112921
PGPUB-FILING-TYPE: new
DOCUMENT-IDENTIFIER: US 20030112921 A1

TITLE: Methods and devices for analysis of x-ray images

PUBLICATION-DATE: June 19, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Lang, Philipp	Lexington	MA	US	
Steines, Daniel	Palo Alto	CA	US	

US-CL-CURRENT: 378/54

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 13. Document ID: US 20030109951 A1

L13: Entry 13 of 70

File: PGPB

Jun 12, 2003

PGPUB-DOCUMENT-NUMBER: 20030109951

PGPUB-FILING-TYPE: new

DOCUMENT-IDENTIFIER: US 20030109951 A1

TITLE: Monitoring system for an industrial process using one or more multidimensional variables

PUBLICATION-DATE: June 12, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Hsiung, Chang-Meng B.	Irvine	CA	US	
Munoz, Bethsabeth	Pasadena	CA	US	
Roy, Ajoy Kumar	Pasadena	CA	US	
Steinthal, Michael Gregory	Los Angeles	CA	US	
Sunshine, Steven A.	Pasadena	CA	US	
Vicic, Michael Allen	Pasadena	CA	US	
Zhang, Shou-Hua	Arcadia	CA	US	

US-CL-CURRENT: 700/108; 700/116, 700/117, 700/96

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 14. Document ID: US 20030104499 A1

L13: Entry 14 of 70

File: PGPB

Jun 5, 2003

PGPUB-DOCUMENT-NUMBER: 20030104499

PGPUB-FILING-TYPE: new

DOCUMENT-IDENTIFIER: US 20030104499 A1

TITLE: Cell-based detection and differentiation of lung cancer

PUBLICATION-DATE: June 5, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Pressman, Norman J.	Glencoe	IL	US	
Hirsch, Kenneth S.	Redwood City	CA	US	
Hirsch, Adrian	Redwood City	CA	US	

US-CL-CURRENT: 435/7.23; 435/287.2, 435/366

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

RMIC	Draw Desc	Image
------	-----------	-------

☐ 15. Document ID: US 20030083756 A1

L13: Entry 15 of 70

File: PGPB

May 1, 2003

PGPUB-DOCUMENT-NUMBER: 20030083756

PGPUB-FILING-TYPE: new

DOCUMENT-IDENTIFIER: US 20030083756 A1

TITLE: Temporary expanding integrated monitoring network

PUBLICATION-DATE: May 1, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Hsiung, Chang-Meng B.	Irvine	CA	US	
Munoz, Bethsabeth	Pasadena	CA	US	
Roy, Ajoy Kumar	Pasadena	CA	US	
Steinthal, Michael Gregory	Los Angeles	CA	US	
Sunshine, Steven A.	Pasadena	CA	US	
Vicic, Michael Allen	Pasadena	CA	US	
Zhang, Shou-Hua	Arcadia	CA	US	

US-CL-CURRENT: 700/28

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

RMIC	Draw Desc	Image
------	-----------	-------

☐ 16. Document ID: US 20030069877 A1

L13: Entry 16 of 70

File: PGPB

Apr 10, 2003

PGPUB-DOCUMENT-NUMBER: 20030069877

PGPUB-FILING-TYPE: new

DOCUMENT-IDENTIFIER: US 20030069877 A1

TITLE: System for automatically generating queries

PUBLICATION-DATE: April 10, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Grefenstette, Gregory T.	Gieres	PA	FR	
Shanahan, James G.	Pittsburgh		US	

US-CL-CURRENT: 707/2

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

RMIC	Draw Desc	Image
------	-----------	-------

☐ 17. Document ID: US 20030065409 A1

L13: Entry 17 of 70

File: PGPB

Apr 3, 2003

PGPUB-DOCUMENT-NUMBER: 20030065409
PGPUB-FILING-TYPE: new
DOCUMENT-IDENTIFIER: US 20030065409 A1

TITLE: Adaptively detecting an event of interest

PUBLICATION-DATE: April 3, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Raeth, Peter G.	Beavercreek	OH	US	
Bostick, Randall L.	Springboro	OH	US	
Bertke, Donald Allen	Beavercreek	OH	US	

US-CL-CURRENT: 700/31; 700/28, 700/30, 700/44

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 18. Document ID: US 20030061201 A1

L13: Entry 18 of 70

File: PGPB

Mar 27, 2003

PGPUB-DOCUMENT-NUMBER: 20030061201
PGPUB-FILING-TYPE: new
DOCUMENT-IDENTIFIER: US 20030061201 A1

TITLE: System for propagating enrichment between documents

PUBLICATION-DATE: March 27, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Grefenstette, Gregory T.	Gieres	PA	FR	
Shanahan, James G.	Pittsburgh		US	

US-CL-CURRENT: 707/3

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 19. Document ID: US 20030061200 A1

L13: Entry 19 of 70

File: PGPB

Mar 27, 2003

PGPUB-DOCUMENT-NUMBER: 20030061200
PGPUB-FILING-TYPE: new
DOCUMENT-IDENTIFIER: US 20030061200 A1

TITLE: System with user directed enrichment and import/export control

PUBLICATION-DATE: March 27, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Hubert, Laurence	St Bernard du Touvet		FR	
Guerin, Nicolas	Grenoble		FR	

US-CL-CURRENT: 707/3

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 20. Document ID: US 20030051026 A1

L13: Entry 20 of 70

File: PGPB

Mar 13, 2003

PGPUB-DOCUMENT-NUMBER: 20030051026

PGPUB-FILING-TYPE: new

DOCUMENT-IDENTIFIER: US 20030051026 A1

TITLE: Network surveillance and security system

PUBLICATION-DATE: March 13, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Carter, Ernst B.	San Francisco	CA	US	
Zolotov, Vasily	San Francisco	CA	US	

US-CL-CURRENT: 709/224; 706/909, 713/201

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 21. Document ID: US 20030046421 A1

L13: Entry 21 of 70

File: PGPB

Mar 6, 2003

PGPUB-DOCUMENT-NUMBER: 20030046421

PGPUB-FILING-TYPE: new

DOCUMENT-IDENTIFIER: US 20030046421 A1

TITLE: Controls and displays for acquiring preferences, inspecting behavior, and guiding the learning and decision policies of an adaptive communications prioritization and routing system

PUBLICATION-DATE: March 6, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Horvitz, Eric J.	Kirkland	WA	US	
Baribault, Gregory P.	Lynnwood	WA	US	

US-CL-CURRENT: 709/238; 709/206

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 22. Document ID: US 20030033347 A1

L13: Entry 22 of 70

File: PGPB

Feb 13, 2003

PGPUB-DOCUMENT-NUMBER: 20030033347

PGPUB-FILING-TYPE: new
DOCUMENT-IDENTIFIER: US 20030033347 A1

TITLE: Method and apparatus for inducing classifiers for multimedia based on unified representation of features reflecting disparate modalities

PUBLICATION-DATE: February 13, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Bolle, Rudolf M.	Bedford Hills	NY	US	
Haas, Norman	Mount Kisco	NY	US	
Oles, Frank J.	Peekskill	NY	US	
Zhang, Tong	Tuckahoe	NY	US	

US-CL-CURRENT: 709/107; 709/100, 709/201

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 23. Document ID: US 20030033288 A1

L13: Entry 23 of 70

File: PGPB

Feb 13, 2003

PGPUB-DOCUMENT-NUMBER: 20030033288
PGPUB-FILING-TYPE: new
DOCUMENT-IDENTIFIER: US 20030033288 A1

TITLE: Document-centric system with auto-completion and auto-correction

PUBLICATION-DATE: February 13, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Shanahan, James G.	Pittsburgh	PA	US	
Grefenstette, Gregory T.	Gieres		FR	

US-CL-CURRENT: 707/3

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 24. Document ID: US 20030033287 A1

L13: Entry 24 of 70

File: PGPB

Feb 13, 2003

PGPUB-DOCUMENT-NUMBER: 20030033287
PGPUB-FILING-TYPE: new
DOCUMENT-IDENTIFIER: US 20030033287 A1

TITLE: Meta-document management system with user definable personalities

PUBLICATION-DATE: February 13, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Shanahan, James G.	Pittsburgh	PA	US	
Grefenstette, Gregory T.	Gieres		FR	
Fernstrom, Christer	St-Ismier		FR	
Hubert, Laurence	St Bernard du Touvet		FR	

US-CL-CURRENT: 707/3

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments	KWIC	Draw Desc	Image
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------	------	-----------	-------

☐ 25. Document ID: US 20030033145 A1

L13: Entry 25 of 70

File: PGPB

Feb 13, 2003

PGPUB-DOCUMENT-NUMBER: 20030033145
PGPUB-FILING-TYPE: new
DOCUMENT-IDENTIFIER: US 20030033145 A1

TITLE: System, method, and article of manufacture for detecting emotion in voice signals by utilizing statistics for voice signal parameters

PUBLICATION-DATE: February 13, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Petrushin, Valery A.	Arlington Heights	IL	US	

US-CL-CURRENT: 704/236

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments	KWIC	Draw Desc	Image
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------	------	-----------	-------

☐ 26. Document ID: US 20030023444 A1

L13: Entry 26 of 70

File: PGPB

Jan 30, 2003

PGPUB-DOCUMENT-NUMBER: 20030023444
PGPUB-FILING-TYPE: new
DOCUMENT-IDENTIFIER: US 20030023444 A1

TITLE: A VOICE RECOGNITION SYSTEM FOR NAVIGATING ON THE INTERNET

PUBLICATION-DATE: January 30, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
ST. JOHN, VICKI	BOULDER CREEK	CA	US	

US-CL-CURRENT: 704/270.1

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments	KWIC	Draw Desc	Image
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------	------	-----------	-------

☐ 27. Document ID: US 20030004966 A1

L13: Entry 27 of 70

File: PGPB

Jan 2, 2003

PGPUB-DOCUMENT-NUMBER: 20030004966
PGPUB-FILING-TYPE: new
DOCUMENT-IDENTIFIER: US 20030004966 A1

TITLE: Business method and apparatus for employing induced multimedia classifiers
based on unified representation of features reflecting disparate modalities

PUBLICATION-DATE: January 2, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Bolle, Rudolf M.	Bedford Hills	NY	US	
Haas, Norman	Mount Kisco	NY	US	
Oles, Frank J.	Peekskill	NY	US	
Zhang, Tong	Tuckahoe	NY	US	

US-CL-CURRENT: 707/104.1

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 28. Document ID: US 20020194247 A1

L13: Entry 28 of 70

File: PGPB

Dec 19, 2002

PGPUB-DOCUMENT-NUMBER: 20020194247
PGPUB-FILING-TYPE: new
DOCUMENT-IDENTIFIER: US 20020194247 A1

TITLE: Method and system for scheduling execution of activities

PUBLICATION-DATE: December 19, 2002

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Caggese, Sergio	Rome		IT	
MacLellan, Scot	Rome		IT	

US-CL-CURRENT: 709/102

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 29. Document ID: US 20020183977 A1

L13: Entry 29 of 70

File: PGPB

Dec 5, 2002

PGPUB-DOCUMENT-NUMBER: 20020183977
PGPUB-FILING-TYPE: new
DOCUMENT-IDENTIFIER: US 20020183977 A1

TITLE: Endpoint detection in substrate fabrication processes

PUBLICATION-DATE: December 5, 2002

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Sui, Zhifeng	Milpitas	CA	US	
Luscher, Paul E.	Sunnyvale	CA	US	
Johansson, Nils	Los Gatos	CA	US	
Welch, Michael D.	Livermore	CA	US	

US-CL-CURRENT: 702/188

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 30. Document ID: US 20020169562 A1

L13: Entry 30 of 70

File: PGPB

Nov 14, 2002

PGPUB-DOCUMENT-NUMBER: 20020169562

PGPUB-FILING-TYPE: new

DOCUMENT-IDENTIFIER: US 20020169562 A1

TITLE: Defining biological states and related genes, proteins and patterns

PUBLICATION-DATE: November 14, 2002

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Stephanopoulos, Gregory	Chester	MA	US	
Misra, Jatin	Cambridge	MA	US	
Hwang, Daehee	Cambridge	MA	US	
Schmitt, William A. JR.	Boston	MA	US	
Alevizos, Ilias	Watertown	MA	US	
Silva, Saliya Sudharshana	Kandy	CO	LK	
Gill, Ryan T.	Boulde		US	

US-CL-CURRENT: 702/19; 435/6, 530/350, 536/23.1

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 31. Document ID: US 20020168664 A1

L13: Entry 31 of 70

File: PGPB

Nov 14, 2002

PGPUB-DOCUMENT-NUMBER: 20020168664

PGPUB-FILING-TYPE: new

DOCUMENT-IDENTIFIER: US 20020168664 A1

TITLE: Automated pathway recognition system

PUBLICATION-DATE: November 14, 2002

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Murray, Joseph	Berkeley	CA	US	
Hendrix, Donna	Berkeley	CA	US	
Chin, Daniel J.	Foster City	CA	US	

US-CL-CURRENT: 435/6; 702/20

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 32. Document ID: US 20020151992 A1

L13: Entry 32 of 70

File: PGPB

Oct 17, 2002

PGPUB-DOCUMENT-NUMBER: 20020151992

PGPUB-FILING-TYPE: new

DOCUMENT-IDENTIFIER: US 20020151992 A1

TITLE: Media recording device with packet data interface

PUBLICATION-DATE: October 17, 2002

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Hoffberg, Steven M.	West Harrison	NY	US	
Hoffberg-Borghesani, Linda I.	Acton	MA	US	

US-CL-CURRENT: 700/83; 700/17, 700/18, 700/86, 700/87

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 33. Document ID: US 20020114425 A1

L13: Entry 33 of 70

File: PGPB

Aug 22, 2002

PGPUB-DOCUMENT-NUMBER: 20020114425

PGPUB-FILING-TYPE: new

DOCUMENT-IDENTIFIER: US 20020114425 A1

TITLE: Methods and devices for analysis of X-ray images

PUBLICATION-DATE: August 22, 2002

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Lang, Philipp	Lexington	MA	US	
Steines, Daniel	Palo Alto	CA	US	

US-CL-CURRENT: 378/56

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 34. Document ID: US 20020107843 A1

L13: Entry 34 of 70

File: PGPB

Aug 8, 2002

PGPUB-DOCUMENT-NUMBER: 20020107843

PGPUB-FILING-TYPE: new

DOCUMENT-IDENTIFIER: US 20020107843 A1

TITLE: Customer self service subsystem for classifying user contexts

PUBLICATION-DATE: August 8, 2002

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Biebesheimer, Debra L.	Carmel	NY	US	
Keller, Neal M.	Somers	NY	US	
Oblinger, Daniel A.	New York	NY	US	
Podlaseck, Mark E.	New Preston	CT	US	
Rolando, Stephen J.	Katonah	NY	US	

US-CL-CURRENT: 707/3

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 35. Document ID: US 20020095259 A1

L13: Entry 35 of 70

File: PGPB

Jul 18, 2002

PGPUB-DOCUMENT-NUMBER: 20020095259

PGPUB-FILING-TYPE: new

DOCUMENT-IDENTIFIER: US 20020095259 A1

TITLE: Multiparameter analysis for drug response and related methods

PUBLICATION-DATE: July 18, 2002

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Hood, Leroy E.	Seattle	WA	US	
Siegel, Andrew F.	Shoreline	WA	US	

US-CL-CURRENT: 702/19; 435/7.1

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 36. Document ID: US 20020087649 A1

L13: Entry 36 of 70

File: PGPB

Jul 4, 2002

PGPUB-DOCUMENT-NUMBER: 20020087649

PGPUB-FILING-TYPE: new

DOCUMENT-IDENTIFIER: US 20020087649 A1

TITLE: Bounded-deferral policies for reducing the disruptiveness of notifications

PUBLICATION-DATE: July 4, 2002

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Horvitz, Eric J.	Kirkland	WA	US	

US-CL-CURRENT: 709/207

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 37. Document ID: US 20020083067 A1

L13: Entry 37 of 70

File: PGPB

Jun 27, 2002

PGPUB-DOCUMENT-NUMBER: 20020083067

PGPUB-FILING-TYPE: new

DOCUMENT-IDENTIFIER: US 20020083067 A1

TITLE: Enterprise web mining system and method

PUBLICATION-DATE: June 27, 2002

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
Tamayo, Pablo	Cambridge	MA	US	
Myczkowski, Jacek	Harvard	MA	US	
Campos, Marcos	Cambridge	MA	US	

US-CL-CURRENT: 707/100

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 38. Document ID: US 20020010587 A1

L13: Entry 38 of 70

File: PGPB

Jan 24, 2002

PGPUB-DOCUMENT-NUMBER: 20020010587

PGPUB-FILING-TYPE: new

DOCUMENT-IDENTIFIER: US 20020010587 A1

TITLE: SYSTEM, METHOD AND ARTICLE OF MANUFACTURE FOR A VOICE ANALYSIS SYSTEM THAT DETECTS NERVOUSNESS FOR PREVENTING FRAUD

PUBLICATION-DATE: January 24, 2002

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
PERTRUSHIN, VALERY A.	ARLINGTON HEIGHTS	IL	US	

US-CL-CURRENT: 704/275

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 39. Document ID: US 20020002464 A1

L13: Entry 39 of 70

File: PGPB

Jan 3, 2002

PGPUB-DOCUMENT-NUMBER: 20020002464

PGPUB-FILING-TYPE: new

DOCUMENT-IDENTIFIER: US 20020002464 A1

TITLE: SYSTEM AND METHOD FOR A TELEPHONIC EMOTION DETECTION THAT PROVIDES OPERATOR FEEDBACK

PUBLICATION-DATE: January 3, 2002

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
PETRUSHIN, VALERY A.	ARLINGTON HEIGHTS	IL	US	

US-CL-CURRENT: 704/275

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 40. Document ID: US 20020002460 A1

L13: Entry 40 of 70

File: PGPB

Jan 3, 2002

PGPUB-DOCUMENT-NUMBER: 20020002460

PGPUB-FILING-TYPE: new

DOCUMENT-IDENTIFIER: US 20020002460 A1

TITLE: SYSTEM METHOD AND ARTICLE OF MANUFACTURE FOR A VOICE MESSAGING EXPERT SYSTEM THAT ORGANIZES VOICE MESSAGES BASED ON DETECTED EMOTIONS

PUBLICATION-DATE: January 3, 2002

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
PERTRUSHIN, VALERY	ARLINGTON HEIGHTS	IL	US	

US-CL-CURRENT: 704/270

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 41. Document ID: US 20010056349 A1

L13: Entry 41 of 70

File: PGPB

Dec 27, 2001

PGPUB-DOCUMENT-NUMBER: 20010056349

PGPUB-FILING-TYPE: new

DOCUMENT-IDENTIFIER: US 20010056349 A1

TITLE: 69VOICE AUTHENTICATION SYSTEM AND METHOD FOR REGULATING BORDER CROSSING

PUBLICATION-DATE: December 27, 2001

INVENTOR-INFORMATION:

NAME	CITY	STATE	COUNTRY	RULE-47
ST. JOHN, VICKI	BOULDER CREEK	CA	US	

US-CL-CURRENT: 704/270

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 42. Document ID: US 6640145 B2

L13: Entry 42 of 70

File: USPT

Oct 28, 2003

US-PAT-NO: 6640145

DOCUMENT-IDENTIFIER: US 6640145 B2

TITLE: Media recording device with packet data interface

DATE-ISSUED: October 28, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Hoffberg; Steven	West Harrison	NY	10604	
Hoffberg-Borghesani; Linda	Acton	MA	01720	

US-CL-CURRENT: 700/83; 700/17, 700/19, 700/23, 704/200, 704/201, 704/7, 709/200,
709/201, 709/202

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 43. Document ID: US 6629081 B1

L13: Entry 43 of 70

File: USPT

Sep 30, 2003

US-PAT-NO: 6629081

DOCUMENT-IDENTIFIER: US 6629081 B1

TITLE: Account settlement and financing in an e-commerce environment

DATE-ISSUED: September 30, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Cornelius; Richard D.	Santa Monica	CA		
Stepniczka; Andreas	San Francisco	CA		
Chu; Kevin	Atlanta	GA		

US-CL-CURRENT: 705/30

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 44. Document ID: US 6606744 B1

L13: Entry 44 of 70

File: USPT

Aug 12, 2003

US-PAT-NO: 6606744

DOCUMENT-IDENTIFIER: US 6606744 B1

TITLE: Providing collaborative installation management in a network-based supply chain environment

DATE-ISSUED: August 12, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Mikurak; Michael G.	Hamilton	NJ		

US-CL-CURRENT: 717/174; 705/26, 717/178

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 45. Document ID: US 6574754 B1

L13: Entry 45 of 70

File: USPT

Jun 3, 2003

US-PAT-NO: 6574754

DOCUMENT-IDENTIFIER: US 6574754 B1

TITLE: Self-monitoring storage device using neural networks

DATE-ISSUED: June 3, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Smith; Gordon James	Rochester	MN		

US-CL-CURRENT: 714/47; 706/26, 709/224, 714/26, 714/42

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 46. Document ID: US 6556977 B1

L13: Entry 46 of 70

File: USPT

Apr 29, 2003

US-PAT-NO: 6556977

DOCUMENT-IDENTIFIER: US 6556977 B1

TITLE: Methods for selecting, developing and improving diagnostic tests for pregnancy-related conditions

DATE-ISSUED: April 29, 2003

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Lapointe; Jerome	Oakland	CA		
DeSieno; Duane D.	La Jolla	CA		

US-CL-CURRENT: 706/15; 706/23, 706/45

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 47. Document ID: US 6480194 B1

L13: Entry 47 of 70

File: USPT

Nov 12, 2002

US-PAT-NO: 6480194

DOCUMENT-IDENTIFIER: US 6480194 B1

TITLE: Computer-related method, system, and program product for controlling data visualization in external dimension(s)

DATE-ISSUED: November 12, 2002

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Sang'udi; Gerald P.	Sunnyvale	CA		
Bott; Ross A.	Half Moon Bay	CA		
Tesler; Joel D.	Cupertino	CA		
Hawkes; John R.	Mercer Island	WA		
Xiong; Rebecca W.	Cambridge	MA		
Schkolnick; Mario	Mountain View	CA		

US-CL-CURRENT: 345/440; 345/473

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 48. Document ID: US 6460049 B1

L13: Entry 48 of 70

File: USPT

Oct 1, 2002

US-PAT-NO: 6460049

DOCUMENT-IDENTIFIER: US 6460049 B1

TITLE: Method system and computer program product for visualizing an evidence classifier

DATE-ISSUED: October 1, 2002

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Becker; Barry G.	Mountain View	CA		
Kohavi; Ron	Mountain View	CA		
Sommerfield; Daniel A.	San Mateo	CA		
Tesler; Joel D.	Cupertino	CA		

US-CL-CURRENT: 707/104.1; 707/100, 707/101, 707/102

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 49. Document ID: US 6453246 B1

L13: Entry 49 of 70

File: USPT

Sep 17, 2002

US-PAT-NO: 6453246

DOCUMENT-IDENTIFIER: US 6453246 B1

TITLE: System, method, and computer program product for representing proximity data in a multi-dimensional space

DATE-ISSUED: September 17, 2002

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Agrafiotis; Dimitris K.	Downingtown	PA		
Labanov; Victor S.	Yardley	PA		
Salemme; Francis R.	Yardley	PA		

US-CL-CURRENT: 702/27; 382/225, 702/179, 702/32

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 50. Document ID: US 6421612 B1

L13: Entry 50 of 70

File: USPT

Jul 16, 2002

US-PAT-NO: 6421612

DOCUMENT-IDENTIFIER: US 6421612 B1

**** See image for Certificate of Correction ****

TITLE: System, method and computer program product for identifying chemical compounds having desired properties

DATE-ISSUED: July 16, 2002

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Agrafiotis; Dimitris K.	Downingtown	PA		
Bone; Roger F.	Bridgewater	NJ		
Salemme; Francis R.	Yardley	PA		
Soll; Richard M.	Lawrenceville	NJ		

US-CL-CURRENT: 702/19

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

[Generate Collection](#)[Print](#)

Terms	Documents
L12 and (bayesian or neural adj network)	70

Display Format:[Change Format](#)[Previous Page](#)[Next Page](#)

WEST[Generate Collection](#)[Print](#)**Search Results - Record(s) 51 through 70 of 70 returned.**☐ 51. Document ID: US 6418424 B1

L13: Entry 51 of 70

File: USPT

Jul 9, 2002

US-PAT-NO: 6418424

DOCUMENT-IDENTIFIER: US 6418424 B1

TITLE: Ergonomic man-machine interface incorporating adaptive pattern recognition based control system

DATE-ISSUED: July 9, 2002

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Hoffberg; Steven M.	West Harrison	NY	10604	
Hoffberg-Borghesani; Linda I.	Acton	MA	01720	

US-CL-CURRENT: 706/21; 434/178, 706/52[Full](#) [Title](#) [Citation](#) [Front](#) [Review](#) [Classification](#) [Date](#) [Reference](#) [Sequences](#) [Attachments](#)[KWC](#) [Draw Desc](#) [Image](#)☐ 52. Document ID: US 6400996 B1

L13: Entry 52 of 70

File: USPT

Jun 4, 2002

US-PAT-NO: 6400996

DOCUMENT-IDENTIFIER: US 6400996 B1

TITLE: Adaptive pattern recognition based control system and method

DATE-ISSUED: June 4, 2002

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Hoffberg; Steven M.	West Harrison	NY	10994	
Hoffberg-Borghesani; Linda I.	Acton	MA	01720	

US-CL-CURRENT: 700/83; 345/810, 345/840, 345/841, 370/218, 370/355, 700/17, 700/24, 700/25, 700/86, 700/87, 709/102, 709/223, 709/227, 709/318[Full](#) [Title](#) [Citation](#) [Front](#) [Review](#) [Classification](#) [Date](#) [Reference](#) [Sequences](#) [Attachments](#)[KWC](#) [Draw Desc](#) [Image](#)☐ 53. Document ID: US 6373483 B1

L13: Entry 53 of 70

File: USPT

Apr 16, 2002

US-PAT-NO: 6373483

DOCUMENT-IDENTIFIER: US 6373483 B1

TITLE: Method, system and computer program product for visually approximating scattered data using color to represent values of a categorical variable

DATE-ISSUED: April 16, 2002

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Becker; Barry Glenn	Mountain View	CA		
Crawfis; Roger A.	Dublin	OH		

US-CL-CURRENT: 345/419

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 54. Document ID: US 6353810 B1

L13: Entry 54 of 70

File: USPT

Mar 5, 2002

US-PAT-NO: 6353810

DOCUMENT-IDENTIFIER: US 6353810 B1

**** See image for Certificate of Correction ****

TITLE: System, method and article of manufacture for an emotion detection system improving emotion recognition

DATE-ISSUED: March 5, 2002

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Petrushin; Valery A.	Arlington Heights	IL		

US-CL-CURRENT: 704/236; 704/246, 704/270

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 55. Document ID: US 6336109 B1

L13: Entry 55 of 70

File: USPT

Jan 1, 2002

US-PAT-NO: 6336109

DOCUMENT-IDENTIFIER: US 6336109 B1

TITLE: Method and apparatus for inducing rules from data classifiers

DATE-ISSUED: January 1, 2002

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Howard; Gary	Sawbridgeworth			GB

US-CL-CURRENT: 706/25; 706/12, 706/13, 706/2

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 56. Document ID: US 6301579 B1

L13: Entry 56 of 70

File: USPT

Oct 9, 2001

US-PAT-NO: 6301579

DOCUMENT-IDENTIFIER: US 6301579 B1

TITLE: Method, system, and computer program product for visualizing a data structure

DATE-ISSUED: October 9, 2001

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Becker; Barry G.	Mountain View	CA		

US-CL-CURRENT: 707/102; 345/440, 707/104.1

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 57. Document ID: US 6295514 B1

L13: Entry 57 of 70

File: USPT

Sep 25, 2001

US-PAT-NO: 6295514

DOCUMENT-IDENTIFIER: US 6295514 B1

TITLE: Method, system, and computer program product for representing similarity/dissimilarity between chemical compounds

DATE-ISSUED: September 25, 2001

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Agrafiotis; Dimitris K.	Downington	PA		
Lobanov; Victor S.	Yardley	PA		
Salemme; Francis R.	Yardley	PA		

US-CL-CURRENT: 703/12; 702/179, 702/180, 702/181, 702/27, 702/28, 702/30, 702/31, 702/32, 703/2

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 58. Document ID: US 6278464 B1

L13: Entry 58 of 70

File: USPT

Aug 21, 2001

US-PAT-NO: 6278464

DOCUMENT-IDENTIFIER: US 6278464 B1

TITLE: Method, system, and computer program product for visualizing a decision-tree classifier

DATE-ISSUED: August 21, 2001

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Kohavi; Ron	Mountain View	CA		
Tesler; Joel D.	Cupertino	CA		

US-CL-CURRENT: 345/440

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 59. Document ID: US 6275806 B1

L13: Entry 59 of 70

File: USPT

Aug 14, 2001

US-PAT-NO: 6275806

DOCUMENT-IDENTIFIER: US 6275806 B1

TITLE: System method and article of manufacture for detecting emotion in voice signals by utilizing statistics for voice signal parameters

DATE-ISSUED: August 14, 2001

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Pertrushin; Valery A.	Arlington Heights	IL		

US-CL-CURRENT: 704/272; 704/270

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 60. Document ID: US 6259451 B1

L13: Entry 60 of 70

File: USPT

Jul 10, 2001

US-PAT-NO: 6259451

DOCUMENT-IDENTIFIER: US 6259451 B1

TITLE: Method, system, and computer program product for mapping between an overview and a partial hierarchy

DATE-ISSUED: July 10, 2001

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Tesler; Joel D.	Cupertino	CA		

US-CL-CURRENT: 345/419

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 61. Document ID: US 6151571 A

L13: Entry 61 of 70

File: USPT

Nov 21, 2000

US-PAT-NO: 6151571

DOCUMENT-IDENTIFIER: US 6151571 A

TITLE: System, method and article of manufacture for detecting emotion in voice signals through analysis of a plurality of voice signal parameters

DATE-ISSUED: November 21, 2000

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Pertrushin; Valery A.	Arlington Heights	IL		

US-CL-CURRENT: 704/209; 704/207, 704/270

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 62. Document ID: US 6137499 A

L13: Entry 62 of 70

File: USPT

Oct 24, 2000

US-PAT-NO: 6137499

DOCUMENT-IDENTIFIER: US 6137499 A

TITLE: Method, system, and computer program product for visualizing data using partial hierarchies

DATE-ISSUED: October 24, 2000

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Tesler; Joel D.	Cupertino	CA		

US-CL-CURRENT: 345/440

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 63. Document ID: US 6081750 A

L13: Entry 63 of 70

File: USPT

Jun 27, 2000

US-PAT-NO: 6081750

DOCUMENT-IDENTIFIER: US 6081750 A

TITLE: Ergonomic man-machine interface incorporating adaptive pattern recognition based control system

DATE-ISSUED: June 27, 2000

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Hoffberg; Steven Mark	Yonkers	NY	10701-1705	
Hoffberg-Borghesani; Linda Irene	Acton	MA	01720	

US-CL-CURRENT: 700/17; 345/520, 700/11, 700/56, 700/83, 700/86

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 64. Document ID: US 6034697 A

L13: Entry 64 of 70

File: USPT

Mar 7, 2000

US-PAT-NO: 6034697

DOCUMENT-IDENTIFIER: US 6034697 A

TITLE: Interpolation between relational tables for purposes of animating a data visualization

DATE-ISSUED: March 7, 2000

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Becker; Barry Glenn	Mountain View	CA		

US-CL-CURRENT: 345/606

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 65. Document ID: US 5960435 A

L13: Entry 65 of 70

File: USPT

Sep 28, 1999

US-PAT-NO: 5960435

DOCUMENT-IDENTIFIER: US 5960435 A

TITLE: Method, system, and computer program product for computing histogram aggregations

DATE-ISSUED: September 28, 1999

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Rathmann; Peter K.	San Mateo	CA		
Haber; Eben M.	Palo Alto	CA		

US-CL-CURRENT: 707/101; 707/102

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 66. Document ID: US 5930803 A

L13: Entry 66 of 70

File: USPT

Jul 27, 1999

US-PAT-NO: 5930803

DOCUMENT-IDENTIFIER: US 5930803 A

**** See image for Certificate of Correction ****

TITLE: Method, system, and computer program product for visualizing an evidence

classifier

DATE-ISSUED: July 27, 1999

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Becker; Barry G.	Mountain View	CA		
Kohavi; Ron	Mountain View	CA		
Sommerfield; Daniel A.	San Mateo	CA		
Tesler; Joel D.	Cupertino	CA		

US-CL-CURRENT: 707/104.1; 345/804, 707/100

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 67. Document ID: US 5901246 A

L13: Entry 67 of 70

File: USPT

May 4, 1999

US-PAT-NO: 5901246

DOCUMENT-IDENTIFIER: US 5901246 A

TITLE: Ergonomic man-machine interface incorporating adaptive pattern recognition based control system

DATE-ISSUED: May 4, 1999

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Hoffberg; Steven M.	Yonkers	NY	10701-1705	
Hoffberg-Borghesani; Linda I.	Acton	MA	01720	

US-CL-CURRENT: 382/209

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 68. Document ID: US 5875108 A

L13: Entry 68 of 70

File: USPT

Feb 23, 1999

US-PAT-NO: 5875108

DOCUMENT-IDENTIFIER: US 5875108 A

TITLE: Ergonomic man-machine interface incorporating adaptive pattern recognition based control system

DATE-ISSUED: February 23, 1999

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Hoffberg; Steven M.	Yonkers	NY	10701-1705	
Hoffberg-Borghesani; Linda I.	Acton	MA	01720	

US-CL-CURRENT: 700/17; 382/181, 382/190, 700/83

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 69. Document ID: US 5861891 A

L13: Entry 69 of 70

File: USPT

Jan 19, 1999

US-PAT-NO: 5861891

DOCUMENT-IDENTIFIER: US 5861891 A

**** See image for Certificate of Correction ****

TITLE: Method, system, and computer program for visually approximating scattered data

DATE-ISSUED: January 19, 1999

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Becker; Barry Glenn	Mountain View	CA		

US-CL-CURRENT: 345/619

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

☐ 70. Document ID: US 5701400 A

L13: Entry 70 of 70

File: USPT

Dec 23, 1997

US-PAT-NO: 5701400

DOCUMENT-IDENTIFIER: US 5701400 A

TITLE: Method and apparatus for applying if-then-else rules to data sets in a relational data base and generating from the results of application of said rules a database of diagnostics linked to said data sets to aid executive analysis of financial data

DATE-ISSUED: December 23, 1997

INVENTOR-INFORMATION:

NAME	CITY	STATE	ZIP CODE	COUNTRY
Amado; Carlos Armando	Miami	FL	33131-2400	

US-CL-CURRENT: 706/45; 706/47, 706/60

Full	Title	Citation	Front	Review	Classification	Date	Reference	Sequences	Attachments
------	-------	----------	-------	--------	----------------	------	-----------	-----------	-------------

KWIC	Draw Desc	Image
------	-----------	-------

Generate Collection

Print

Terms	Documents
L12 and (bayesian or neural adj network)	70

Display Format:

-

Change Format

[Previous Page](#)

[Next Page](#)

Learning to Classify Text from Labeled and Unlabeled Documents (1998) (Make Corrections) (57 citations)

Kamal Nigam

Proceedings of AAAI-98, 15th Conference of the American Association for Artificial Intelligence

View or download:

cmu.edu/~mccallum/...mcataai98s.ps.gz

umass.edu/~mccallu...mcataai98s.ps.gz

Cached: [PS.gz](#) [PS](#) [PDF](#) [DjVu](#) [Image](#) [Update](#) [Help](#)

From: cmu.edu/afs/cs/project/j...node30 (more)

Homepages: [K.Nigam](#) [HPSearch](#) [\(Update Links\)](#)

CiteSeer
Electronic Literature Digital Library

[Home/Search](#) [Context](#) [Related](#)

[\(Enter summary\)](#)

Rate this article: 1 2 3 4 5 (best)
[Comment on this article](#)

Abstract: This paper shows that the accuracy of learned **text** classifiers can be improved by augmenting small numbers of labeled training documents with a large pool of unlabeled documents. This is significant because in many important **text** classification problems obtaining classification labels is expensive, while large quantities of unlabeled documents are readily available. We present a theoretical argument showing that, under common assumptions, unlabeled data contain information about the target...
[\(Update\)](#)

Context of citations to this paper: [More](#)

...information is available; in these cases, we would like to have some way of using that information to improve our model. Nigam et al. [10] studied the problem of learning to classify data into pre existing categories in the presence of labeled and unlabeled examples. Their...

...relevance feedback, using unlabeled documents that seem relevant to modify its believed relevant document distribution. Nigam et al. [20] got good results with this approach using the multinomial model, and we conjecture that our model, more accurate than the multinomial,...

Cited by: [More](#)

Using Artificial Anomalies to Detect Unknown and Known.. - Fan, Miller, Stolfo (2001) [\(Correct\)](#)

Mining Newsgroups Using Networks Arising From Social.. - Agrawal, Rajagopalan.. [\(Correct\)](#)

Submitted (3/98) to the 15th International Conference.. - Improving **Text**.. [\(Correct\)](#)

Similar documents (at the sentence level):

47.3%: Using EM to Classify **Text** from Labeled and Unlabeled Documents - Nigam (1998) [\(Correct\)](#)

27.5%: Learning to Classify **Text** from Labeled and Unlabeled.. - Nigam, McCallum, Thrun, .. (1998) [\(Correct\)](#)

10.1%: **Text** Classification from Labeled and Unlabeled.. - Nigam, McCallum.. (1999) [\(Correct\)](#)

Active bibliography (related documents): [More](#) [All](#)

0.6: Employing EM and Pool-Based Active Learning for **Text** Classification - McCallum (1998) [\(Correct\)](#)

0.3: An Adaptive Web Page Recommendation Service - Balabanovic (1997) [\(Correct\)](#)

0.3: Book Recommending Using **Text** Categorization with Extracted.. - Mooney (1998) [\(Correct\)](#)

Similar documents based on text: [More](#) [All](#)

0.5: A Parallel Learning Algorithm for **Text** Classification - Kruengkrai, Jaruskulchai [\(Correct\)](#)

0.5: Using Unlabeled Data to Improve **Text** Classification - Nigam (2001) [\(Correct\)](#)

0.4: Pool-Based Active Learning for **Text** Classification - Nigam, McCallum (1998) [\(Correct\)](#)

Related documents from co-citation: [More](#) [All](#)

18: **Text** categorization with Support Vector Machines: Learning with many relevant fe.. - Joachims - 1998

18: Machine Learning (context) - Mitchell - 1997 [Book Details from Amazon or Barnes & Noble](#)

15: Combining labeled and unlabeled data with co-training - Blum, Mitchell - 1998

BibTeX entry: [\(Update\)](#)

K. Nigam, A. McCallum, S. Thrun, and T. Mitchell. Learning to classify **text** from labeled and unlabeled documents. In Proceedings of the Fifteenth National Conference on Artificial Intelligence. AAAI Press, 1998.
<http://citeseer.nj.nec.com/nigam98learning.html> [More](#)

```
@inproceedings{ nigam98learning,  
  author = "Kamal Nigam and Andrew K. McCallum and Sebastian Thrun and Tom M. Mitc  
  title = "Learning to classify text from labeled and unlabeled documents",  
  booktitle = "Proceedings of {AAAI}-98, 15th Conference of the American Associati  
  publisher = "AAAI Press, Menlo Park, US",
```

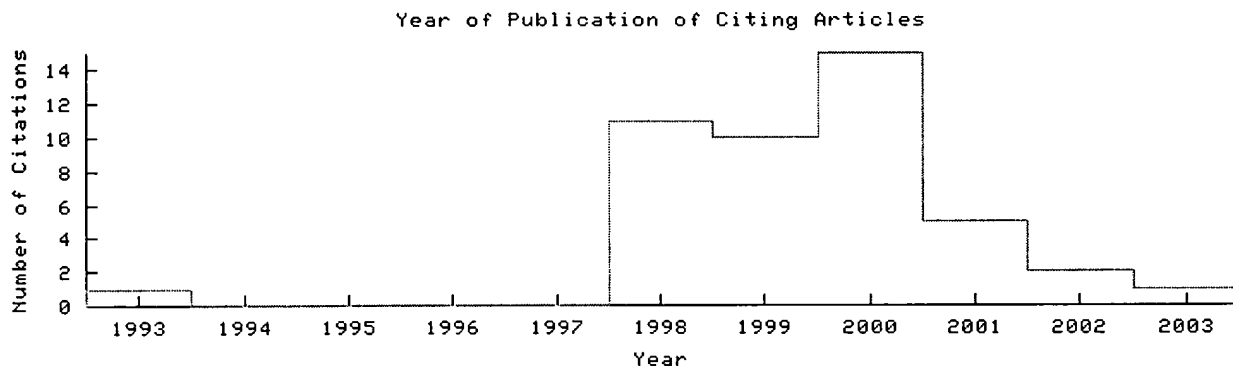
```

address = "Madison, US",
pages = "792--799",
year = "1998",
url = "citeseer.nj.nec.com/nigam98learning.html" }

```

Citations (may not include all citations):

- 1735 Maximum likelihood from incomplete data via the EM (context) - Dempster, Laird et al. - 1977
 221 In Advances in Neural Information Processing Systems (context) - Neural, Systems et al. **Book Details from Amazon or Barnes & Noble**
 116 Learning to extract symbolic knowledge from the World Wide W.. - Craven, Freitag et al. - 1998
 97 Newsweeder: Learning to filter netnews - Lang - 1995
 87 Beyond independence: Conditions for the optimality of the si.. - Domingos, Pazzani - 1997
 82 Bayesian classification (context) - Cheeseman, Stutz - 1996
 77 Context-sensitive learning methods for **text** categorization - Cohen, Singer - 1997
 76 A comparison of two learning algorithms for **text** categorizat.. - Lewis, Ringuette - 1994
 65 Developments in automatic **text** retrieval (context) - Salton - 1991
 61 Supervised learning from incomplete data via an EM approach - Ghahramani, Jordan - 1994
 28 Active learning with committees for **text** categorization (context) - Tadepalli - 1997
 21 A mixture of experts classifier with learning based on both .. (context) - Miller, Uyar - 1997
 18 The effect of unlabeled samples in reducing the small sample.. (context) - Shahshahani, Landgrebe - 1994
 17 Threading electronic mail: A preliminary study - Lewis, Knowles - 1997
 9 An application of least squares fit mapping to **text** informat.. (context) - Yang, Chute - 1993
 8 and Billsus (context) - Pazzani, Muramatsu - 1996
 2 A sequential algorithm for training **text** classifiers (context) - Conference, Learning et al. - 1994
 1 **Text** categorization with Support Vector Machines: Learning w.. (context) - Conference, Learning et al.
 1 Improving retrieval performance by relevance feedback (context) - Hall, -- et al. - 1990
 1 Estimations of dependences based on statistical data (context) - Trans, Remote et al. - 1982



The graph only includes citing articles where the year of publication is known.

Documents on the same site (<http://www.cs.cmu.edu/afs/cs/project/jair/pub/volume8/fuernkranz98a-html/node30.html>): [More](#)

- Conditions for Occam's Razor Applicability and Noise Elimination - Gamberger, Lavrac (1997) [\(Correct\)](#)
 Tractable Induction and Classification in First Order .. - Michèle.. (1997) [\(Correct\)](#)
 Query by Committee - Seung, Oppen, Sompolinsky (1992) [\(Correct\)](#)

[Online articles have much greater impact](#) [More about CiteSeer](#) [Add search form to your site](#) [Submit documents](#) [Feedback](#)

CiteSeer - citeseer.org - [Terms of Service](#) - [Privacy Policy](#) - Copyright © 1997-2002 [NEC Research Institute](#)

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.